

Birla Central Library

PILANI (Jaipur State)

Engg College Branch

Class No :- 621.384132

Book No :- 17538T

Accession No :- 32535

Acc. No

ISSUE LABEL

Not later than the latest date stamped below.

29 Nov 28

EX Feb 5

**THEORY OF
GASEOUS CONDUCTION
AND ELECTRONICS**

*The quality of the materials used in the manufacture
of this book is governed by continued postwar shortages.*

THEORY OF GASEOUS CONDUCTION AND ELECTRONICS

BY

FREDERICK A. MAXFIELD, PH.D.

*U.S. Naval Ordnance Laboratory
on Leave from the University of Wisconsin*

AND

R. RALPH BENEDICT, M.S.

*Associate Professor of Electrical Engineering
University of Wisconsin*

FIRST EDITION
THIRD IMPRESSION

McGRAW-HILL BOOK COMPANY, INC.

NEW YORK AND LONDON

1941

THEORY OF GASEOUS CONDUCTION AND ELECTRONICS

COPYRIGHT, 1941, BY THE
MCGRAW-HILL BOOK COMPANY, INC.

PRINTED IN THE UNITED STATES OF AMERICA

*All rights reserved. This book, or
parts thereof, may not be reproduced
in any form without permission of
the publishers.*

To

EDWARD BENNETT

PREFACE

The need for a textbook covering the fundamentals of conduction in gases and suitable for undergraduate students in electrical engineering and applied physics has been felt for some time, inasmuch as gaseous conduction processes are playing an increasingly important part in the operation of devices and apparatus employed in engineering work today. It is the feeling of the authors that the average student graduating in electrical engineering or applied physics should not only be familiar with the fundamental theory of operation of high vacuum electronic equipment but should also have some knowledge of the theory of gaseous conduction. The importance of this is emphasized by a consideration of the wide diversity of specialized fields into which the students will go after graduation. These range all the way from work in radio and communications through industrial work involving the operation or design of apparatus and control equipment to work in the field of high-voltage engineering, power transmission, and central stations. Yet in all these the students will find applications of the theory of gaseous conduction. To satisfy this need, the book discusses not only high-vacuum conduction as found in radio tubes, cathode-ray tubes, and phototubes, but also the theory and application of corona, sparking, glows, and arcs.

Because it was felt that a broad fundamental course of this type was desirable as a background and preparation for the specialized studies of the senior year, a one-semester course covering most of the material of this book has been required since 1935 of all junior students in electrical engineering at the University of Wisconsin. Mimeographed notes were written for this course, which after additions and revisions have evolved into this book.

In writing this we have tried to keep in mind that a statement of the scientific principles involved in electronic and gaseous conduction is much more valuable than a detailed description of specific apparatus and applications, especially since the students have such diverse interests and will have an

opportunity in future studies to pursue these special interests with greater facility, having had a fundamental course of this nature.

We have, accordingly, endeavored to cover the essential principles underlying the phenomena of gaseous conduction and electronics rather than to provide an exhaustive treatment of the subject. As far as possible we have made the treatment quantitative rather than purely descriptive. Examples of the application of gaseous conduction drawn from engineering practice are included, but no detailed treatment of devices and circuit applications is attempted. We feel that by placing the emphasis on the theory of operation of gaseous conduction devices, their inherent characteristics, useful features, and limitations will be better understood.

Where questions of a controversial nature have been encountered, some attempt has been made to indicate the rival theories, but sometimes it has been necessary to give without elaboration what seems to the authors to be the best interpretation. The references provided will indicate avenues for further study and sources of more complete bibliographies.

The primary purpose of the book is thus to interpret in a systematic way the underlying phenomena upon which the properties of *all types* of gaseous conduction devices depend. This implies a rather limited study of specific devices and their applications, but a more extensive knowledge of the characteristics and applications of specific devices can readily be obtained by further study after the fundamental principles governing the operation of these devices have been mastered.

We wish to acknowledge our indebtedness to the authors of all the books and technical papers that have been used as source material and references for this book. In addition we wish to thank the Allis-Chalmers Mfg. Co., General Electric Co., Ohio Brass Co., RCA Mfg. Co., and the Westinghouse Electric and Mfg. Co. for furnishing the material for a number of the figures used in this book.

We are particularly indebted to Prof. Edward Bennett for encouraging us to undertake this task and to carry it to completion, and to Dr. Arthur E. Ruark for his constant encouragement and his many helpful suggestions for the revision of the manuscript.

The authors' wives, Mary De Weese Maxfield and Dorothy Stauss Benedict, have been a constant inspiration and help both in the preparation of the manuscript and in the reading of proof.

The authors wish to extend their thanks to all the student NYA workers and others who aided in the preparation of the manuscript and illustrations.

The manuscript was prepared while both authors were associated with the Electrical Engineering Department of the University of Wisconsin, but one was called for service in the Naval Ordnance Laboratory before the proof was received. The authors wish to express their indebtedness to the University of Wisconsin and to its Electrical Engineering Department for the opportunity to write this book.

FREDERICK A. MAXFIELD,
R. RALPH BENEDICT.

WASHINGTON, D. C.,
MADISON, WIS.,
August, 1941.

CONTENTS

	Page
PREFACE.	vii
CHAPTER	
I. INTRODUCTION	1
1-1. Scope of Gaseous Conduction and Electronics	1
1-2. Mechanism of Conduction.	2
1-3. Metallic and Gaseous Conduction.	5
1-4. Applications	9
1-5. Object of This Book	11
1-6. Units	12
 II. THE ELECTROSTATIC FIELD.	17
2-1. Introduction.	17
2-2. Electric Intensity at a Point	19
2-3. Electric Potential.	23
2-4. Relation between Potential and Electric Intensity	27
2-5. Gauss's Theorem	29
2-6. Mapping Electric Fields.	37
2-7. The Method of Images	40
 III. MOVEMENT OF CHARGED PARTICLES IN HIGH VACUUM.	45
3-1. Forces in the Electric Field.	45
3-2. Laws of Motion in an Electrostatic Field.	46
3-3. Laws of Motion in a Uniform Magnetic Field.	51
3-4. The Cathode-ray Oscillograph: the Electrostatic Deflect- ing Type.	54
3-5. The Magnetic Deflecting Type Oscillograph	60
3-6. Effect of Relativity Mass Correction	62
3-7. Application of the Oscillograph.	65
3-8. Focusing of Electron Beams	68
3-9. Anode Rays	71
3-10. Television	74
 IV. THE MOLECULAR THEORY OF GASES.	81
4-1. Introduction.	81
4-2. Experimental Basis of Gas Theory: The General Gas Law for Ideal Gases.	82
4-3. The Kinetic Picture of a Perfect Gas	84

CHAPTER	PAGE
4-4. Elementary Calculation of Pressure and Molecular Speed.	85
4-5. Avogadro's Hypothesis	88
4-6. Temperature and Molecular Energy and Speed.	90
4-7. Mixed Gases.	91
4-8. Distribution of Molecular Velocities: Introduction	93
4-9. Maxwell's Deduction of the Velocity Distribution Laws . .	97
4-10. Representation of Maxwell's Distribution Laws.	102
4-11. The Root-mean-square and Arithmetical Average Speeds	106
4-12. Free Paths of Molecules and Electrons.	109
4-13. Law of Distribution of Free Paths.	114
4-14. Calculation of the Number of Molecules Crossing a Plane Surface	118
4-15. Evaporation of Electrons from Metals, or Thermionic Emission.	120
V. CONDUCTION PROCESSES	125
5-1. Conductivity in the Normal State.	125
5-2. Ionizing and Deionizing Agents.	126
VI. ELECTRON EMISSION: THEORY AND TYPICAL APPLICATIONS. . .	135
6-1. Thermionic Emission: Qualitative Observations, and Theories.	135
6-2. Theory of Electrons in Metals and of the Thermionic Emission of Electrons.	138
6-3. Temperature-limited Emission	147
6-4. Electron Emission from Composite Surfaces: Practical Emitters.	149
6-5. Contact Difference of Potential.	160
6-6. The Schottky Effect. Field Emission.	162
6-7. Space-charge-limited or -controlled Electron Currents . .	167
6-8. Typical Applications of the High-vacuum Diode: Recti- fiers; Voltmeters	174
6-9. Grid Control of Electron Space-current: The High- vacuum Triode.	177
6-10. Secondary-electron Emission Caused by Electron Impact	186
6-11. Multigrid Thermionic Tubes; Applications of Grid- controlled Tubes	191
6-12. Photoelectric Electron Emission	206
6-13. Photoelectric Tubes and Their Application.	213
6-14. Secondary Electron Emission Caused by Positive-ion Impact	220
VII. DEIONIZING PROCESSES	225
7-1. Electric Field—Mobility.	225
7-2. Diffusion.	235
7-3. Recombination in the Gas	244
7-4. Recombination at Walls and Electrodes	254

CONTENTS

xiii

CHAPTER	PAGE
VIII. PRODUCTION OF IONS IN THE GAS	259
8-1. Ionization Energies and Photoelectric Ionization in the Gas	259
8-2. Thermal Ionization	263
8-3. Ionization by Collisions of Electrons with Gas Atoms	265
8-4. Ionization by Collisions of Positive Ions with Gas Atoms	268
8-5. The Townsend Discharge	270
8-6. The Ionizing Coefficients α and β	279
IX. SPARKING POTENTIAL, SPARK-OVER, AND CORONA DISCHARGE	287
9-1. Types of Discharge	287
9-2. Sparking Potential in a Uniform Field. Paschen's Law	290
9-3. Sparking Intensity in a Uniform Field.	296
9-4. Sparking in Vacuum, in Gases under High Pressure, and in Mixed Gases.	298
9-5. Sparking, Corona, and Breakdown in Nonuniform Fields	301
9-6. Sparking and Corona between Concentric Cylinders and between Parallel Wires in Air.	305
9-7. Corona Loss on Transmission Lines	308
X. THE GLOW DISCHARGE	311
10-1. Introduction.	311
10-2. The Cathode, or Crookes, Dark Space.	316
10-3. The Negative Glow and the Faraday Dark Space.	325
10-4. The Uniform Positive Column	327
10-5. The Striated Positive Column	334
10-6. Determination of Positive-column Characteristics.	337
10-7. The Anode in the Glow	344
10-8. Stability of Discharges.	345
10-9. Applications of the Glow Discharge.	348
XI. THE ARC DISCHARGE	357
11-1. Introduction.	357
11-2. Total Arc Characteristics	363
11-3. The Arc Positive Column	368
11-4. The Cathode of the Arc	379
11-5. Dynamic Arc Characteristics.	386
11-6. The Arc Oscillator	388
11-7. Alternating-current Arc Characteristics	393
11-8. Extinction of Arcs	397
11-9. Circuit Interrupters.	403
11-10. Mercury-arc Rectifiers.	414
11-11. Arc-backs in Mercury-arc Rectifiers	420
11-12. Thermionic Gas Diode: The Phanotron	424
11-13. Thermionic Gas Triode: The Thyatron	431
11-14. Thyatron Applications	436
11-15. The Ignitron.	443

CHAPTER	PAGE
11-16. Inverters.	446
11-17. Gaseous-discharge Lamps	450
APPENDICES	461
I. List of Symbols.	461
II. Important Constants	464
III. Molecular Weights, and Molecular Speeds and Mean Free Paths at Standard Conditions	465
INDEX.	467

THEORY OF GASEOUS CONDUCTION AND ELECTRONICS

CHAPTER I

INTRODUCTION

1-1. Scope of Gaseous Conduction and Electronics.—All gases ordinarily are good insulators. Under normal circumstances, then, we should not expect appreciable conduction to take place between two separated conductors in a gas. The fact that there are conditions under which a gas conducts electricity readily implies that charge carriers appear in the gas and constitute the medium of conduction. These charge carriers will be designated as *ions*, and any process by which these charge carriers are made to appear will be called an *ionizing process*. It must be remembered that an ion, as we have defined it, is not one kind of charged particle, but any charged particle regardless of its mass or the sign of its electric charge. An ion may have a mass essentially the same as that of the gas molecules, or it may have a mass large enough to suggest that it is made up of several gas molecules.¹ On the other hand, it may have a mass as small as $1/1,840$ times the mass of the hydrogen atom. These small ions are generally called electrons or positrons, depending upon the charge carried by the ion. Electrons are negatively charged particles, and positrons are positively charged particles. In general, the only ions with which we shall be concerned in this work will be *electrons*, *positive molecular ions*, and *negative molecular ions*. Positrons will not be discussed since they are

¹ Whether an ion consists of a single molecule or a group of molecules depends principally upon the gas pressure, the size of the ion increasing with the pressure. At atmospheric pressure, most of the molecular ions still consist of a single molecule.

infrequently encountered in the study of gaseous conduction, and ions compounded out of a group of molecules will be mentioned in only a few instances.

When the conduction of electricity between two separated conductors is due primarily to the motion of electrons, the process can be called *electronic conduction*. This is the type of conduction that is encountered in the ordinary radio tube and in the high-voltage kenotron rectifier. When more than one kind of charge carrier is present, however, the more general term *gaseous conduction* is usually applied. Thus, when a switch is opened in a d.c. circuit in air at atmospheric pressure or when two carbon electrodes, connected to the terminals of a generator, are separated, the intense electric discharge that results is due to the motion of positive and negative molecular ions as well as to electrons that are formed in the gas between the electrodes. Gaseous conduction is important, even at low pressures. For example, in the mercury-arc rectifier the arc chamber is pumped to a relatively high vacuum; yet it is possible to maintain an arc between the mercury pool and another electrode some distance away. The process involved is again gaseous conduction, but here the gas is principally mercury vapor evaporated from the mercury surface by the heat generated in the arc itself.

Since, as we shall see later, the conduction processes are so similar and since a general term is often necessary, we shall frequently speak of *gaseous conduction* and include in this designation all classes of discharges at all pressures, even those which obviously depend simply upon electronic conduction. However, we shall reserve the term electronic conduction for those discharges in which electrons are the only charge carriers. This classification of discharges will perhaps not seem quite so arbitrary when we consider that even with the highest vacuums attainable today there are still more than 10,000 million gas molecules present in every cubic inch of the gas space and positive and negative molecular ions still contribute a slight amount to the total current.

1-2. Mechanism of Conduction.—Current density in a gas can be defined as a vector quantity having the same direction as the direction of motion of the charged particles in the gas and a magnitude equal to the rate of transfer of positive charge through the gas per unit area perpendicular to the direction of motion:

of the charged particles. We can then say that if we have ions of only one kind present

$$\mathbf{J} = Nq\bar{\mathbf{V}} \quad (1-1)$$

Here N represents the number of ions present in a unit volume¹ of the gas or the concentration of ions, q is the positive charge carried by each ion, and $\bar{\mathbf{V}}$ is the average velocity with which the ions travel. Thus, Nq gives the quantity of charge in a unit volume, and $Nq\bar{\mathbf{V}}$ gives the rate of transfer of charge per unit area in the direction of $\bar{\mathbf{V}}$, or the current density. $\mathbf{J}$ is a vector quantity² directed in the same direction as $\bar{\mathbf{V}}$ if q is positive and opposite to $\bar{\mathbf{V}}$ if q is negative.

In general, ions of more than one kind are present in any particular discharge, and in this case the current density is given by Eq. (1-2).

$$\mathbf{J} = N_1q_1\bar{\mathbf{V}}_1 + N_2q_2\bar{\mathbf{V}}_2 + \cdots \quad (1-2)$$

Here, N , q , and $\bar{\mathbf{V}}$ have the same meaning as before, and the subscripts refer to the various kinds of ions present. That is, N_1 , q_1 , and $\bar{\mathbf{V}}_1$ refer to the concentration, charge, and average velocity of the ions of type 1; N_2 , q_2 , and $\bar{\mathbf{V}}_2$ are the corresponding quantities for the ions of type 2; etc. The charges q_1 , q_2 , etc., may be either positive or negative quantities according to the sign of the particular kind of ion concerned. We shall assume as a further simplification here and throughout this book that the magnitude of the charge carried by all ions is equal to the magnitude of the charge carried by the electron. The negative charge carried by the electron (-1.6×10^{-19} coulomb) is the smallest quantity of charge it is possible to obtain. In general, all ions carry a positive or a negative charge that is some integral multiple of the electronic charge, but we shall discuss only singly charged ions.

In the particular case shown in Fig. 1-1, we have positive molecular ions, negative molecular ions, and electrons present in a discharge tube, and the charge carried by each ion is just equal to the electronic charge q_e . Furthermore, all the ions move either to the right or to the left as indicated by the arrows, and there is no motion in the other coordinate directions. If we

¹ See Sec. 1-6 for a discussion of the units used.

² All vector quantities in this book are designated by boldface letters.

assign the subscript 1 to the positive molecular ions, the subscript 2 to the negative molecular ions, and the subscript 3 to the electrons, q_1 will be a positive quantity, q_2 will be a negative

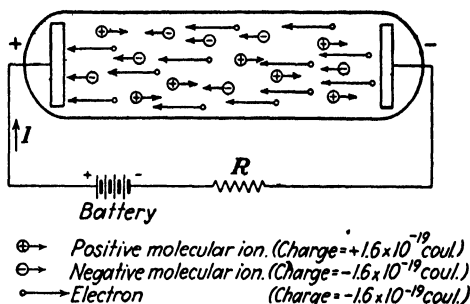


FIG. 1-1.

quantity, and q_3 will be a negative quantity. Also $\bar{V}_2$ and $\bar{V}_3$ will be directed opposite to $\bar{V}_1$ since the positive charges are attracted to the negative plate and the negative charges to the positive plate. If we designate directions to the right as positive, the resultant sign of all terms will thus be positive; the total current density will be the sum of three positive quantities and will be directed to the right as shown by the equations below¹ and in Fig. 1-2.

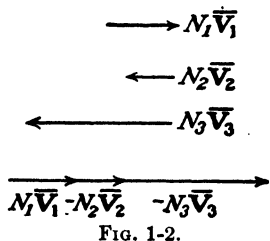


FIG. 1-2.

$$\begin{aligned}
 J &= N_1 \bar{V}_1 \times 1.6 \times 10^{-19} - N_2 \bar{V}_2 \times 1.6 \times 10^{-19} - N_3 \bar{V}_3 \times 1.6 \\
 &\quad \times 10^{-19} \\
 &= 1.6 \times 10^{-19} (N_1 \bar{V}_1 - N_2 \bar{V}_2 - N_3 \bar{V}_3) \\
 &= 1.6 \times 10^{-19} (N_1 \bar{V}_1 + N_2 \bar{V}_2 + N_3 \bar{V}_3) \text{ (directed to the right)}
 \end{aligned}$$

Equation (1-2) tells us that, in order to determine the current which will flow through a gas on the application of a voltage, we must first determine N and $\bar{V}$. These quantities depend on the applied potential and the electric-field distribution; but, in turn, the electric-field and potential distribution generally depend on N and $\bar{V}$. The relations that develop are complicated; but, in

¹ The absolute value of a vector quantity is designated throughout this book by the same symbol as the vector but in ordinary rather than in bold-face letters.

principle at least, the relations can be determined, the current can be expressed as a function of the applied voltage, and problems dealing with gaseous conduction can be solved.

1-3. Metallic and Gaseous Conduction.—Problems involving conduction in gases are not so simple as those involving metallic conduction. There is no simple relation between current and voltage comparable with Ohm's law which can be applied to all gaseous conductors. For the case of *metallic* conduction. Ohm's law can be put in the form of Eq. (1-3).

$$\mathbf{J} = \frac{1}{\rho} \mathbf{F} \quad (1-3)$$

In this equation, $\mathbf{J}$ is the current density, ρ is the resistivity of the medium under study, and $\mathbf{F}$ is the electric intensity at the point at which $\mathbf{J}$ is desired. It is easy to see that this equation follows directly from Eq. (1-2), if we adopt the usual picture for the process of conduction in a metal. In a metal the atoms are packed very closely together and are held in a state of equilibrium by the electrostatic field due to adjacent atoms. Since each atom consists of a positive nucleus and one or more electrons confined to certain orbits about the nucleus, it is possible for the outermost, or least tightly bound, electrons to exist in more or less a state of equilibrium between the closely packed atoms. That is, the atoms of the metal are conceived to be so closely packed that the orbits of the outer electrons overlap. An electron on such an orbit, then, could easily be transferred from atom to atom in the metal. It is these electrons which constitute the so-called "free electrons."

Equation (1-2) can be applied to the calculation of the current in a metal.

$$\mathbf{J} = N_1 q_1 \bar{\mathbf{V}}_1 + N_2 q_2 \bar{\mathbf{V}}_2 \quad (1-2a)$$

The positive ions in the metal are bound together and are free to execute only small vibrations about their equilibrium point; hence, they will not contribute to the current density. "Free electrons" are the only negative ions; hence the expression for $\mathbf{J}$ will contain only one term.

$$\mathbf{J} = Nq\bar{\mathbf{V}} \quad (1-2b)$$

Nq , or the charge per unit volume in the metal, is directly proportional to the conductivity or inversely proportional to the

resistivity of the metal. The average drift velocity of the electrons, $\bar{V}$, is directly proportional to the applied electric intensity. If the proportionality constants are combined with the resistivity, we see that Eq. (1-3) follows directly from Eq. (1-2b):

$$\mathbf{J} = \frac{1}{\rho} \mathbf{E} \quad (1-3)$$

The more familiar form of Ohm's law applying to homogeneous metals maintained at constant temperature can be derived from Eq. (1-3). The electric intensity in a homogeneous metallic conductor of constant cross-sectional area is equal to the voltage applied divided by the length of the conductor. The resistivity ρ in this conductor is equal to the total resistance R multiplied by the cross-sectional area of the conductor and divided by its length. And the current density $\mathbf{J}$ is equal to the total current flowing divided by the cross-sectional area of the conductor. If we make these substitutions for the terms of Eq. (1-3), we see that Eq. (1-4) follows directly.

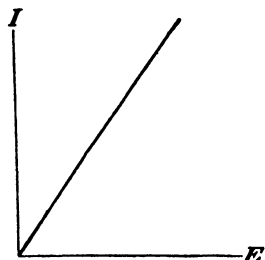


Fig. 1-3.—Metallic conductor.

$$I = \frac{1}{R} E \quad (1-4)$$

By Eq. (1-4) the volt-ampere curve for a homogeneous metallic conductor is a straight line as shown in Fig. 1-3.

In gaseous conductors the relations between E and I are by no means so simple as in the case of metallic conductors. In general, the motions not only of electrons must be considered but those of positive molecular ions and negative molecular ions as well. The concentration of any one kind of charged particle is not constant, and the velocities of the particles depend in a complex way on the applied potential. Experimental curves of the variation of current with impressed voltage are, in particular cases, similar to those shown in Figs. 1-4, 1-5, and 1-6. Curve *a* in Fig. 1-4, for instance, shows that, as the potential is increased on a high-vacuum phototube, the illumination being maintained constant, the current at first increases but soon becomes constant. No matter how much the potential is

increased, the current never exceeds this limiting value until breakdown occurs outside the tube. A similar phototube containing a small amount of inert gas such as helium or argon may at first behave exactly like the high-vacuum phototube as the potential is increased; but, instead of the current remaining constant with increased voltage, a certain voltage may be reached at which the current again begins to increase (curve *b*, Fig. 1-4). This time the current may increase very rapidly with increase in potential and ultimately may tend to approach infinity or a value determined by the external circuit. Other devices have other characteristics: for instance, a hot-cathode

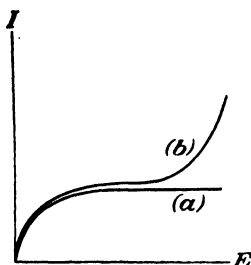


FIG. 1-4.—Phototube.

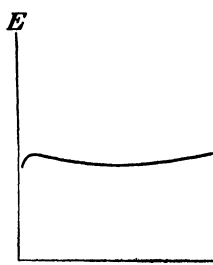


FIG. 1-5.—Hot-cathode rectifier.

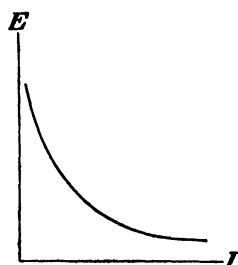


FIG. 1-6.—Carbon arc in air.

mercury-vapor rectifier tube (such as the 866A) has a practically constant voltage drop even though the current through the tube is varied through wide limits by means of the external circuit (Fig. 1-5) and a d.c. arc in air between carbon electrodes has a decreasing voltage drop as the current is increased (Fig. 1-6).

In a metallic conductor it is possible, since the voltage-current characteristic is a straight line, to define the resistance as the ratio of the voltage applied to the current flowing in the conductor, and the number thus determined is a constant. In gaseous conductors, as can be seen from Figs. 1-4, 1-5, and 1-6, this ratio will vary over a wide range as either the current or the voltage of the tube is changed. In some cases, as in a spark gap in air, the resistance, if we mean by resistance the ratio of voltage difference appearing to the current flowing in the discharge, may be millions of ohms at one value of impressed voltage; but if the voltage is raised slightly and breakdown occurs, the resistance may be only a fraction of an ohm. In many discharges the resistance may vary over wide limits in various parts of the

discharge. For instance, in a glow discharge there is a large voltage drop near the cathode, or negative electrode, but only a relatively small voltage drop in the rest of the discharge. For a distance of a few millimeters adjacent to the cathode the resistance may be of the order of 10,000 ohms, but in the rest of the discharge which may be several centimeters long the resistance may be only one-tenth as great.

Still another definition of resistance is obtained if the slope of the $E - I$ curve is taken as the resistance. This may be called a dynamic resistance since it is important when the current or voltage is changing. In the case of metallic conductors, this resistance is the same as the other resistance defined above. Here, again, it can be seen from Figs. 1-4, 1-5, and 1-6 that this dynamic resistance varies over wide limits in any one discharge or from one type of discharge to another. It is even possible to have negative values of dynamic resistance since the arc characteristic (Fig. 1-6) is a falling one.

The fact that gaseous conductors have such a wide variety of characteristics is not an undesirable property of these discharges. The fact that we do have such things as gaseous conductors makes it possible to design such apparatus as the Poulsen arc generator, the operation of which depends on the fact that an arc has a falling volt-ampere characteristic, the mercury-arc rectifier, which has a very low resistance to current flow in one direction and a very high resistance to current flow in the opposite direction, and the radio tube, in which the resistance to current flow can be varied simply by controlling the potential of auxiliary electrodes. But to understand how the characteristics of gaseous conductors can be applied to engineering problems, we shall have to understand something of the mechanism of these discharges and the general principles involved in gaseous conduction. It is not enough to know the characteristics of certain types of tube. If we do not understand the processes involved in the operation of these tubes or if we cannot predict their behavior under special and exceptional circumstances, we cannot fully apply them, in the engineering sense. Of course, we should also like to be able to predict new uses for these gaseous conductors as well as new devices utilizing gaseous conduction. That this is not a barren field for the present-day engineer can best be illustrated by citing a few cases in which rapid development of a gaseous-

conduction device has taken place only in the last few years, although the fundamental principle was discovered a number of years ago.

1-4. Applications.—One of the outstanding pieces of apparatus that has undergone this development is the thermionic vacuum tube. The fact that gases in the neighborhood of red-hot metals become conducting has been known since about 1725 when Du Fay¹ first recorded the observation. However, during the next 150 years following his discovery, the phenomenon was studied only intermittently and no great progress was made in developing an acceptable explanation. Edison, among others, observed the presence of thermionic emission in vacuum in 1883; and soon after that Elster and Geitel, in 1889, demonstrated that the thermionic diode could be used to rectify alternating voltages. But it was not until after the beginning of this century that any great progress was made in explaining or utilizing the properties of thermionic emitters.

In 1902 and 1903, Richardson² investigated thoroughly and gave the first satisfactory explanation of thermionic emission. The next year the first practical use of the diode as a detector of radio signals was made by J. A. Fleming,³ and a few years later Lee De Forest introduced a grid electrode between the thermionic filament and the plate.⁴ The power of this third electrode to control the plate current was soon appreciated, and since that time the development has been rapid. It is interesting to note that, although more than 200 years have elapsed since the discovery of the phenomenon of thermionic emission, it has been only during the last 40 years that any rapid development has taken place; the limit of development is still not in sight, for new uses are constantly being found for tubes employing thermionic emitters.

A similar development could be outlined for the phototube starting with the discovery of Hertz⁵ in 1887 that ultraviolet light shining on a spark gap reduces the voltage necessary for

¹ *Mém. de l'Acad.*, 1733.

² *Proc. Cambridge Phil. Soc.*, **11**, 286, 1902; *Phil. Trans.*, **201**, 516, 1903; *Proc. Roy. Soc. (London)*, **71**, 415, 1903.

³ British Patent 24850, Nov. 16, 1904; U.S. Patent 803684, Apr. 19, 1905.

⁴ U.S. Patent 841387; Jan. 15, 1907.

⁵ *Wied. Ann.*, **31**, 983, 1887.

sparking to occur. Again there was a lapse of a number of years before the newly discovered phenomenon found a practical application. But once engineers realized the possibilities for practical development of these fundamental phenomena the way was opened for a very rapid development and widespread application of the devices conceived in such an apparently, but only apparently, leisurely fashion.

The development of the mercury-arc rectifier follows this same pattern. Using an a.c. generator, in 1882, M. Jamin discovered that if he established an arc between a mercury-pool electrode and a carbon electrode the average value of the current that flowed was not zero as it would have been had he used two similar carbon electrodes.¹ He found, in fact, that the arc acted as a rectifier converting alternating current into pulsating direct current. Other investigators at about the same time were discovering that some rectification of alternating current could be accomplished at reduced air pressure by using two electrodes of quite different size, one large and one small. The discharge that they used was not an arc, but a glow; consequently, the magnitude of the rectified current was quite small.

These experiments were the starting points for the development of arc rectifiers, but we find that about 20 years elapsed before the utility of the mercury-arc rectifier began to be appreciated. Toward the end of the year 1902, Peter Cooper-Hewitt devised what was probably the first polyphase mercury-arc rectifier.² He had constructed this rectifier after a study of the characteristics of the mercury-vapor lamp which he invented the year before (1901). The many advantages of this rectifier over existing types were quickly acknowledged by progressive engineers, and by 1905 Steinmetz³ had described before the American Institute of Electrical Engineers (A.I.E.E.) a two-anode full-wave rectifier installation which had been in service for a year in Schenectady supplying energy for 25 mercury-arc street lamps.

From that time to the present and especially during the 1920's, the development of high-capacity metal-tank mercury-arc rectifiers has been very rapid. Some idea of how rapidly the current and voltage capacity of the metal-tank rectifier has been pushed

¹ *Comptes Rendus*, **94**, pp. 1615-1619, 1882; *Phil. Mag.*, **14**, 154-156, 1882.

² *Electrician*, **50**, 509, 1903.

³ *Trans. A.I.E.E.*, **24**, 371, 1905.

upward in the past few years can be gained when we realize that before 1923 the largest current that could be successfully rectified by a single unit was about 1,000 amp.; yet in the next 10 years this was increased by a factor of 10. During the same period the voltage rating of a single unit was pushed from about 2,500 volts to nearly 50,000 volts. These current and voltage ratings do not, of course, apply to the same units. Here, again, we realize that the limit has not been reached along this line of development; for, with the development of successful high-voltage rectifiers for changing large blocks of a.c. to d.c. power and with the development of high-voltage inverters for changing this power back to alternating current, all the possibilities of d.c. power transmission are introduced.

1-5. Object of This Book.—The more the historical development of electronic- and gaseous-conduction devices is reviewed, the more strongly the realization is brought home that it is only as the significance of the fundamental discoveries has come to be fully appreciated by a larger and larger group of scientists that great strides have been made in the utilization of these fundamental principles. With this in mind, we cannot help feeling that a proper approach to the subject of conduction in gases cannot be made simply by an investigation of the characteristics and properties of particular pieces of apparatus; for, after all, practically all the gaseous-conduction apparatus in use today will be superseded by better and more efficient apparatus in a few years. We realize, however, that the fundamental principles underlying any piece of gaseous-conduction apparatus will not change and that if we are well versed in these fundamental principles we shall be able to keep pace and perhaps aid in setting the pace in the development of new equipment.

In accordance with this ideal, applications of gaseous conduction will occupy a secondary position in the scheme of development of this book. The fundamental principles will be discussed both qualitatively and quantitatively. Wherever possible the authors will strive to give accurate quantitative calculations and derivations. Unfortunately, this cannot be done in all cases. There are many properties of gaseous conduction that can be treated only in a qualitative manner because of the complexity of the problem and the many variables involved. In some cases, there is no completely satisfactory theoretical explanation of a

particular effect. In those cases in which quantitative relations cannot be given, the authors will strive to give the best possible discussion of the subject, stating, wherever possible, the limits of error or the accuracy that can be expected in using a particular derived expression.

The development of the subject will be systematic rather than historical. That is, we shall treat one particular phase of gaseous conduction at a time. First, we shall concern ourselves with the motions and the forces acting upon ions and gas particles individually and as a group. Then, we shall discuss the various processes by which ionization is made to appear in a gas, and then the processes which cause ionization to disappear. In this connection, it seems expedient to illustrate the various processes by a consideration of various electronic tubes such as the diode, triode, and phototube. Next, the processes involved in the formation of sparks will be discussed, and this will be followed by a discussion of the stable glow discharge. Finally, the arc discharge and its properties will be discussed. To show the application of these principles to present-day apparatus, a few illustrations will be given as each subject is developed.

1-6. Units.—Any physical quantity can be measured only by comparison with a similar quantity. A unit quantity is, then, a measure in terms of which all other similar quantities can be expressed. Thus, if we say that a certain length is equal to 12 ft., we imply that we have chosen the foot as the unit measure of length and that the length in question is equal to 12 of these units. Obviously, there are an infinite number of possibilities in choosing a unit of any physical quantity. For instance, lengths can be expressed in terms of inches, feet, yards, meters, centimeters, miles, etc. There is, however, a definite relation between each pair of these units; *i.e.*, one can be expressed in terms of another. This is necessary if different people accustomed to different units are to understand one another.

All the mechanical quantities can be expressed in terms of three *fundamental* units. This is a natural consequence of the fact that there are certain laws or physical relations between all mechanical quantities. It is desirable that these physical relations be expressible by the same relation or equation in any system of units. Thus, we should like to choose units so that the expressions

$$\text{Force} = \text{mass} \times \text{acceleration}$$

or

$$\text{Kinetic energy} = \frac{1}{2} \text{ mass} \times (\text{velocity})^2$$

hold true without numerical factors appearing. Systems of units in which this is true are called "unified" systems, and the units that appear in the equations are the *fundamental* units or the *fundamental derived* units.

The three units ordinarily chosen as *fundamental* for these mechanical relations are length, mass, and time.¹ If we are to extend this system to include relations encountered in the study of electricity and magnetism, a fourth fundamental unit is necessary. And if the system of units is to be extended to cover the whole field of physical quantities, *i.e.*, mechanics, electricity and magnetism, and heat, a fifth fundamental unit is necessary. This last is usually taken as temperature. It is in the choice of the fourth unit that the greatest divergence of opinion has occurred. At the present time, there are no less than 12 comprehensive metric unit systems differing principally in the choice of the electromagnetic fundamental unit although there is some disagreement on the units of length and mass.²

It is desirable in any comprehensive system that is to receive universal acceptance that the fundamental and fundamental derived units be quantities that are *physically conceivable* for the average person. That is, they should be quantities that he can experience in everyday work. They should not be too small or too large and should be convenient for comparison with the standard prototype units preserved in the Bureau of Standards at Washington or at the International Bureau of Weights and Measures at Sèvres, France. For example, units of length that are physically conceivable to everyone are the centimeter and the meter, the first being slightly less than $\frac{1}{2}$ in. and the latter slightly more than 1 yd. The gram, however, is a unit of mass that is too small for a convenient fundamental unit—the gram weight being only about $\frac{1}{16}$ oz. A kilogram, on the other hand, is physically conceivable—a kilogram weight being equal to

¹ This is not necessary, however; for force, velocity, and time might conceivably be chosen as fundamental, as well as other combinations of mechanical quantities.

² ESHBACH, O. W., "Handbook of Engineering Fundamentals," p. 3-16, John Wiley & Sons, Inc., New York, 1936.

2.2 lb. Similarly, the dyne and the erg are much too small for convenient units of force and work. They are just numbers to the average person. But a unit 100,000 times as large as the dyne, the dyne-five or newton, equivalent to about $\frac{1}{4}$ lb. force, is easily conceivable; and a unit 10 million times as large as the erg, the joule, equivalent to the work required to lift a $\frac{3}{4}$ -lb. weight 1 ft., is readily visualized by anyone. We might thus go through the whole list of physical and electrical quantities and try to establish a convenient system of fundamental units.

Although we can easily select examples of convenient units for each physical or electrical quantity, the task of selecting a unified system embodying as many of them as possible is not so easy. An attempt has been made, however, to establish as a universal system one that fulfills as many of these requirements as possible. At its meeting in October, 1935,¹ the International Electrotechnical Commission (I.E.C.) adopted the absolute system of mechanical units proposed by Giorgi in 1904 and decided that the actual substitution of this system for the present international system should take place on Jan. 1, 1940. In this system as adopted in 1935, commonly called the "m.k.s. system" since it defines the fundamental units of length, mass, and time as the meter, kilogram, and second, the fundamental electromagnetic unit was not specified. However, at a later meeting² in June, 1938, the I.E.C. formally accepted the recommendation of its advisory committee that the permeability of free space be taken as the fourth fundamental unit. •

The adoption of the permeability of free space as the fourth fundamental unit does not alter the fact that the unit of current will be the ampere, the unit of voltage the volt, and the unit of charge the coulomb since all these units are related by simple physical laws. This is a desirable characteristic of any universal system. For some time, all practical measurements of voltage, current, resistance, and power have been in terms of volts, amperes, ohms, and watts, respectively; for it has been considered easier and more accurate to make electrical comparisons with standards based on a specific galvanic cell and a standard resistance coil than to make absolute measurements in terms of more fundamental units.

¹ KENNELLY, A. E., *Elec. Eng.*, 54, 1383, 1935.

² KENNELLY, A. E., *Elec. Eng.*, 58, 78, 1939.

There is one other question to be settled by the international committee and that is the question of rationalization. The original c.g.s. electrostatic and electromagnetic systems are unrationalized; *i.e.*, the dielectric constant for free space or the permittivity of free space was taken as unity in the electrostatic c.g.s. system. Although this choice simplified some of the fundamental electrostatic and electromagnetic equations such as the equation for the force between two concentrated electric charges or two concentrated magnetic poles, it complicated other more useful expressions, from an engineer's viewpoint, by introducing a factor 4π . Thus, for instance, the expression for the magnetomotive force (m.m.f.) in a magnetic circuit is expressed in the c.g.s. electromagnetic system by

$$\mathcal{F} = 4\pi \Sigma NI \quad \text{(c.g.s. electromagnetic units)} \quad (1-5)$$

whereas in the rationalized system it takes the form

$$\mathcal{F} = \Sigma NI \quad (1-6)$$

Another example is the expression for the capacitance of a parallel-plate condenser. In the c.g.s. electrostatic system, it takes the form

$$C = \frac{\kappa A}{4\pi s} \quad \text{(c.g.s. electrostatic units)} \quad (1-7)$$

whereas in the rationalized system it becomes

$$C = \frac{\epsilon A}{s} \quad (1-8)$$

Here κ and ϵ are the dielectric constant or permittivity in the two systems. In neither of these cases is there any evidence of circular symmetry or other apparent reason for the introduction of the factor 4π . It must be introduced, however, in the c.g.s. forms of the equations if they are to be consistent with the units adopted. Rationalized units are to be preferred in most cases since, with rationalization, equations expressing rectilinear symmetry are free from the factor 4π whereas equations expressing circular symmetry contain the factor 4π .

In view of all this the rationalized m.k.s. system of units has been adopted for use in this book. This system is identical with the original Giorgi system with the exception that, whereas

Giorgi chose the ohm as the additional fundamental unit to supplement the three fundamental mechanical units, in the discussions that follow the permeability of free space has been chosen as the additional fundamental unit.¹ This will not change many of the derived formulas. The electrical quantities used will still be the familiar volt, ampere, ohm, etc.; but the mechanical units will be expressed in meters, kilograms, joules, etc. The complete list of the m.k.s. units that will be used throughout the book is found in the list of symbols of Appendix I. With but a few exceptions, all equations will be expressed in these units.

It is obvious that, although the rationalized m.k.s. system has been adopted for this book, the form of the equations in any other unified system will be unchanged except for the rationalizing factor. For example, all the equations in this book, excepting those involving the permeability μ , are adaptable to the use of unrationalized c.g.s. electrostatic units by simply substituting $\kappa/4\pi$ for the permittivity ϵ , κ being the dielectric constant.

¹ Permittivity is a *fundamental* derived unit in the m.k.s. system related to permeability by the relation $\mu_0\epsilon_0 = 1/c^2$. In this equation μ_0 and ϵ_0 are, respectively, the permeability and the permittivity of free space and c is the velocity of light in vacuum. For a further discussion of permittivity, see Sec. 2-1.

CHAPTER II

THE ELECTROSTATIC FIELD¹

2-1. Introduction.—In the study of the motions of ions in electrical discharges, we often find that the motions are caused mainly by the electrical forces of attraction and repulsion between the ion and the other charges in the vicinity. These charges may include not only the charges on the electrodes but also the charges in the gas space itself. It is the purpose of this chapter to present the principles that will be found helpful in calculating (1) the forces on ions in electric fields and (2) the work done on an ion when it moves from one point to another in an electric field.

The observation of the forces on charged bodies leads to the definition of an electrostatic field as a region in which charged bodies experience a force by reason of their charge or any region that contains one or more charged bodies.

The quantitative relation giving the force acting on each of two neighboring charged bodies is based upon the experiments of Coulomb and is known as Coulomb's law. The following statement gives the essence of Coulomb's law: *The forces exerted upon each other by two small charged bodies vary directly as the product of the charges and inversely as the square of the distance between their centers. The force also depends on the kind of medium in which the charged bodies are placed.*

These forces act along the line joining the centers of the bodies and are repulsive on similarly charged bodies and attractive on oppositely charged bodies. The bodies must be very small compared with their separation and the medium must be extended and homogeneous if Coulomb's law is to hold accurately. This law may be expressed² in the form of an equation, as follows:

¹ General references for Chap. II will be found on p. 44.

² Historically, Coulomb's law was first expressed in the c.g.s. system in the form $f = Q_1 Q_2 / \kappa r^2$ in which f was measured in dynes, the separation r was measured in centimeters and Q_1 and Q_2 were measured in e.s.u. The con-

$$f = \frac{Q_1 Q_2}{4\pi\epsilon r^2} \quad (2-1)$$

In this expression, r is the distance between the centers of the concentrated charges Q_1 and Q_2 . The constant factor which depends on the medium has been written as $1/4\pi\epsilon$. The constant ϵ is called the permittivity of the medium. The rationalizing factor 4π is included in the denominator in order to avoid 4π factors which would otherwise arise in subsequent derived formulas, as explained in Sec. 1-6.

The coulomb has been defined in such a way that if two small bodies are each charged with one coulomb of electricity and are placed in a vacuum with a distance of one meter between centers the force of repulsion will be 8.9892×10^9 newtons.¹ Therefore the quantity $1/4\pi\epsilon$ in Eq. (2-1) has the value 8.9892×10^9 for free space, and we find that the permittivity of free space has the value 8.85×10^{-12} . The value of the permittivity of free space is used as a standard against which the values of the permittivities of material substances are compared. This comparison is conveniently made by letting ϵ_0 designate the permittivity of free space and by expressing the permittivity ϵ as follows:

$$\epsilon = \epsilon_r \epsilon_0 \quad (2-2)$$

The new quantity ϵ_r is called the relative permittivity of the medium, since it is the ratio of the permittivity of the medium in question to the permittivity of free space. The relative permittivity of a medium is also equal to its relative dielectric constant. The value of ϵ_0 was found to be

$$\epsilon_0 = 8.85 \times 10^{-12} \quad (2-3)$$

The value of ϵ_r is very nearly unity for all gases at moderate

stant of proportionality $1/\kappa$ is the factor that depends on the medium. The constant κ in this system is the dielectric constant of the medium, and it has the value unity in free space.

¹ See general reference 2-2, p. 45. From the very great magnitude of the force of repulsion, it is obvious that the coulomb is a relatively large unit quantity of electricity for use in electrostatic calculations. On the other hand the coulomb is a very convenient unit to use in calculations of current, since the ampere is a flow of 1 coulomb/sec. It is natural to choose the coulomb as the unit charge, even though it is of inconvenient size for electrostatic calculations, for calculations of current are made much more frequently than are calculations of electrostatic fields.

pressures (for air at atmospheric pressure ϵ , has the value 1.00059), and therefore little error is introduced by using the permittivity of free space for the permittivity of gases.

2-2. Electric Intensity at a Point.—One method of describing an electric field is in terms of the force exerted on a very small test charge Q_t introduced into the field. If none of the original charges setting up the field moves under the inductive influence of the test charge, the force $\mathbf{f}$ on the test charge will be directly proportional to the quantity of charge Q_t . The constant ratio of $\mathbf{f}$ to Q_t is defined to be the magnitude of the electric intensity vector $\mathbf{F}$. The vector equation defining both the direction and magnitude of the electric intensity is the following:

$$\mathbf{F} = \lim_{Q_t \rightarrow 0} \frac{\mathbf{f}}{Q_t} \quad (Q_t \text{ is a positive charge}) \quad (2-4)$$

The direction of $\mathbf{F}$ is the same as that of the force on a *positive* test charge placed at the point. When $\mathbf{f}$ is measured in newtons and Q_t in coulombs, $\mathbf{F}$ is measured in units called the volt per meter (see Sec. 2-4).¹

An electric field could be described by determining the electric intensity at a large number of points in the field and then drawing vectors setting out from each test point which show the magnitude and direction of the electric intensity at each of these points. In some situations, it is simpler to describe the variation of the magnitude of the electric intensity with distance by an ordinary Cartesian graph and to describe the direction of the intensity by a written statement. The direction of the intensity in a field can also be represented by a system of lines drawn so that the intensity vectors are tangent to a given line at all points along that line. Such a line is called a *line of electric intensity*. Examples of field maps using lines of electric intensity will be given in Sec. 2-6.

It is obvious that if the electric intensity at a point is known, the force on a charge Q , whether it is an electron, ion, or small concentrated charge, placed at the point can be obtained by the expression

$$\mathbf{f} = Q\mathbf{F} \quad (2-5)$$

¹ In c.g.s. units, $\mathbf{F}$ is expressed in statvolts per centimeter, when $\mathbf{f}$ is expressed in dynes and Q_t is in e.s.u.

under the assumption that Q does not modify the original electric field.

The electric intensity at a point in the field of a single concentrated charge Q can be calculated by applying Coulomb's law. Imagine that the test charge Q_t is placed at the given point P which is a distance r from the center of the concentrated charge. The force exerted on Q_t is given by Coulomb's law [Eq. (2-1)]. Substituting this value of the force into the defining equation (2-4) for F , we find

$$F = \frac{Q}{4\pi\epsilon r^2} \quad (2-6)$$

The direction of F is radially away from Q if Q is a positive charge.

If, for example, Q is positive and equal to 10^{-8} coulomb, the

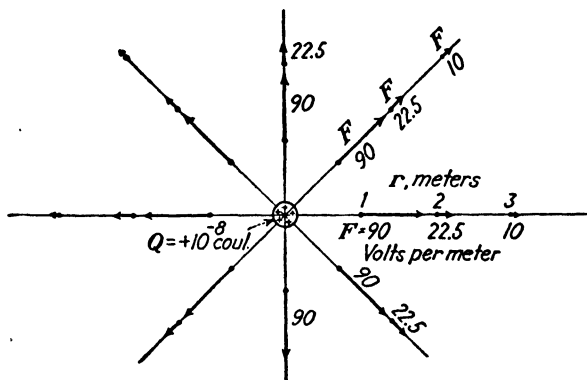


Fig. 2-1.—Intensity vectors in the field of a concentrated charge.

electric-intensity vectors in a plane cutting through the center of the concentrated charge are as shown in Fig. 2-1.

The electric intensity in the field of two or more concentrated charges is obtained by calculating the intensity vectors at the point in question caused by each of the charges in the absence of all the others and then finding the vector sum of all the component intensities. That this sum represents the intensity at the point as given by the defining equation (2-4) is immediately obvious when we reflect that the total force on the test charge would be the vector sum of the forces exerted by the individual charges in the field. This follows from the principle of superposition of electrostatic forces.

As a simple example of the calculation of the intensity in the field of two concentrated charges, we take the situation depicted in Fig. 2-2. The problem is to calculate the electric intensity at the point P located as shown in the figure. The magnitude of the intensity F_1 set up at the point P by the charge Q_1 is

$$F_1 = \frac{Q_1}{4\pi\epsilon r_1^2} = 45 \text{ volts/m.}$$

Similarly the magnitude of the vector F_2 is found to be 18 volts/m. These two vectors must be added vectorially to find the total intensity F . A convenient way to carry out the vector addition analytically is by the method

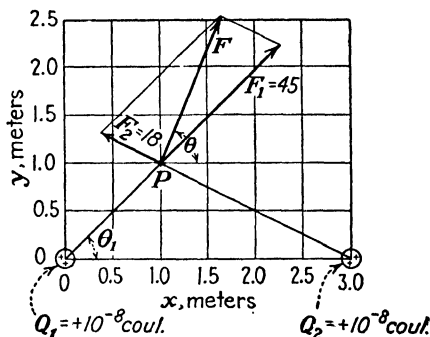


FIG. 2-2.—Electric intensity in the field of two charges.

of components. The component of F_1 parallel to the x -axis in the positive x -direction is $F_1 \cos \theta_1$ which has the value 31.8 volts/m. The component of F_1 parallel to the y -axis is $F_1 \sin \theta_1$, which is also 31.8 volts/m. The components of F_2 parallel to the x -axis and y -axis are -16.1 and $+8.05$ volts/m., respectively. Summing the components along the x - and y -axes, we find

$$\begin{aligned} F_x &= 31.8 - 16.1 = 15.7 \text{ volts/m.} \\ F_y &= 31.8 + 8.05 = 39.85 \text{ volts/m.} \end{aligned}$$

The magnitude of the resultant intensity F is given by

$$F = \sqrt{F_x^2 + F_y^2} = 42.8 \text{ volts/m.}$$

The angle θ that the intensity makes with the x -axis is

$$\theta = \tan^{-1} \left(\frac{F_y}{F_x} \right) = 68.5 \text{ deg.}$$

and now the intensity at the point P has been fully determined.

The intensity set up by a distributed charge is calculated by subdividing the charge into very small elements. The intensities set up by these elements are then added vectorially, the

resultant giving the total intensity. An example will serve to explain the procedure. In Fig. 2-3 is shown a thin glass rod s m. long which is uniformly charged with a charge of $+q$ coulombs/m. of length. The rod is on the x -axis with one end at the origin of coordinates. The problem is to calculate the electric intensity at a point P that is on the y -axis at a distance y from one end of the rod. If the distance y is large compared with the diameter of the rod, the finite diameter of the rod may be neglected, and therefore we are permitted to deal with the charge contained on an element of length dx as if it were a

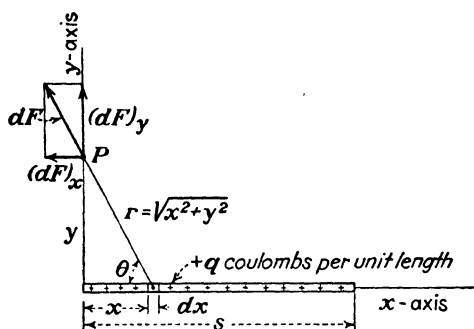


FIG. 2-3.—Electric intensity in the field of a uniformly charged rod.

concentrated charge. The quantity of charge on the length dx is $q dx$. This charge causes an intensity dF at P equal to

$$dF = \frac{q dx}{4\pi\epsilon(y^2 + x^2)^{3/2}}$$

The direction of dF is shown in Fig. 2-3. To add this intensity to those set up by the other elements of the rod, it will again be convenient to use the method of components. The y -component of dF has the value,

$$(dF)_y = dF \sin \theta = \frac{qy dx}{4\pi\epsilon(y^2 + x^2)^{3/2}}$$

After integrating between the limits $x = 0$ and $x = s$, we find

$$F_y = \frac{qs}{4\pi\epsilon y(y^2 + s^2)^{3/2}} \quad (2-7)$$

The value of the x -component is found in a similar way.

$$F_x = \int_{x=0}^{x=s} dF \cos \theta = \int_0^s \frac{qx \, dx}{4\pi\epsilon(y^2 + x^2)^{\frac{3}{2}}} = \frac{q}{4\pi\epsilon} \left[\frac{1}{y} - \frac{1}{(y^2 + s^2)^{\frac{1}{2}}} \right] \quad (2-8)$$

If necessary, these two components can be combined to give the resultant.

2-3. Electric Potential.—An alternative method of describing an electric field is in terms of the work done on a very small test charge as it moves in the field. If the test charge is imagined to be infinitely remote from the charges setting up the field, the electrostatic force acting on the test charge is zero and no work is done on the test charge as it moves. As the test charge approaches the region of the charges, however, the electrostatic forces will come into play, and the test charge will be urged either in the direction of its motion or in the opposite direction, depending on the location and polarities of the charges setting up the field and of the test charge itself. If we consider the test charge to be at rest or moving very slowly in the electrostatic field, there must be some force exerted on the test body by an external agent in order to maintain this equilibrium. This force will be at all times equal and opposite to the electrostatic force. In consequence of the force and displacement, work will be done either *by* or *on* the external agent. To avoid confusion, it is helpful to specify that we shall consider the work done *by* the external agent as positive work and further to specify the use of a *positive* test charge.

If the test charge is so small that it has a negligible inductive effect on the charged bodies in the field, it is found that the work done by the external agent in bringing the test charge up from infinity to a point P in the field is directly proportional to the quantity of the test charge Q_i . Under these conditions, there will be a constant ratio between the work done and the magnitude of the test charge. This constant ratio is defined as the *potential at the point P* . Letting E designate the potential at the point P , we write the definition in the mathematical form

$$E = \frac{W}{Q_i} \quad (Q_i \text{ is a positive charge}) \quad (2-9)$$

The unit of potential is the volt when W is measured in joules and Q_i in coulombs.¹ In Eq. (2-9), W is the algebraic value of the work done *by* the external agent in bringing the very small test charge $+Q_i$ from an infinitely distant region to the point in question. The value of W must be independent of the path along which the test charge is brought to point P if the potential E is to have a definite value. That this is indeed true will be established in the course of the following analysis.

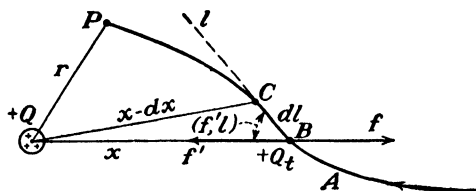


FIG. 2-4.—Test charge Q_t in the field of charge Q .

Let us calculate the potential at a point P that is at a distance r from the center of the concentrated charge $+Q$ as shown in Fig. 2-4. Let the test charge be brought from infinity along the path $ABCP$. The electrostatic force of repulsion will be the radial force f . The “external” force f' is equal and opposite to the electrostatic force. The work dW done by the external agent in moving the test body the distance dl is

$$dW = f' \cos (f', l) dl$$

The projection of dl through the angle (f', l) gives $-dx$. Therefore,

$$dW = -f' dx$$

The force f' has the same magnitude as f ; therefore we apply Coulomb's law to obtain

$$dW = -\frac{QQ_t dx}{4\pi\epsilon x^2}$$

Integrating from ∞ to r ,

$$W = \frac{QQ_t}{4\pi\epsilon r}$$

It is evident that integration along any other path will give the same result, showing that the work done is independent of

¹ In c.g.s. units the unit of potential is the statvolt when W is measured in ergs and Q_i in e.s.u.

the path. After substituting this value of W into the defining equation (2-9), we find the value of the potential at P ,

$$E = \frac{Q}{4\pi\epsilon r} \quad (2-10)$$

Equation (2-10) shows that the potential in the field of a concentrated charge is inversely proportional to the distance from the charge to the point in question. The value of the potential may be either positive or negative depending on the polarity of the charge Q setting up the field and will be measured in volts if Q is in coulombs, r is in meters, and ϵ has the appropriate value given in Sec. 2-1. It is important to note that E depends only on the distance r and the magnitude and polarity of the charge Q setting up the field.

If we again apply the defining equation (2-9) to find the potential at a point in the field of several concentrated charges, we find that the total potential is the algebraic sum of terms of the type in Eq. (2-10), one term for each of the charges setting up the field. Letting these charges be Q_1 , Q_2 , etc., we may express the potential as

$$E = \frac{Q_1}{4\pi\epsilon r_1} + \frac{Q_2}{4\pi\epsilon r_2} + \dots \quad \begin{array}{l} \text{(sum to extend} \\ \text{over all charges)} \end{array} \quad (2-11)$$

In this expression the r 's are the distances from the point P to the charge indicated by the subscripts. This equation shows that the principle of linear superposition applies to the calculation of the potential in the field of a number of charges.

In the most general situation, the charges setting up the field are distributed over the surfaces of the charged bodies in the region and perhaps, also, are distributed through part of the volume of the region. If the charge distribution is known, the potential at any point in the field can be found by an extension of the method of the foregoing paragraph. The charges would be subdivided into small enough elements dQ , so that they could be treated as concentrated charges. Then the potential at a point would be calculated by summing the effect of all the charges in the field. Expressed mathematically, this takes the form

$$E = \sum \frac{dQ}{4\pi\epsilon r} \quad \begin{array}{l} \text{(sum to extend over all} \\ \text{the charges in the field)} \end{array} \quad (2-12)$$

Here r is the distance from dQ to the point P .

As an example of the application of Eq. (2-12) to a distributed charge, let us calculate the potential at the point P in Fig. 2-3. We recall that the charge on the glass rod is uniformly distributed and has the value q coulombs per unit length. As an elementary charge dQ , we take the charge on the length dx , which has the value $q dx$. Substituting in Eq. (2-12) and passing to an integral, we find

$$E = \int_0^s \frac{q dx}{4\pi\epsilon r} = \int_0^s \frac{q dx}{4\pi\epsilon(y^2 + x^2)^{\frac{1}{2}}} = \frac{q}{4\pi\epsilon} \log \left(\frac{s + \sqrt{s^2 + y^2}}{y} \right) \quad (2-13)$$

In the following section, we shall show how the y -component of the electric intensity at point P may be deduced from the expression just derived for the potential at points on the y -axis. This will illustrate one of the useful applications of the concept of potential.

It is often desirable to compare the potentials at two points in a field. For this purpose the *potential increase from a point 1 to a point 2* has been defined to be the algebraic value of the potential at 2 minus that at 1. Thus, letting $\Delta E_{1,2}$ designate the potential increase from 1 to 2, we may write

$$\Delta E_{1,2} = E_2 - E_1 \quad (2-14)$$

In the event that the algebraic sign of the result is unimportant, only the absolute value of ΔE need be considered, and this is called the *potential difference* between the two points.

From the manner in which the potential at a point and the potential increase have been defined, it may be seen that the potential increase from 1 to 2 can be expressed in terms of the work required to move a test charge from 1 to 2. Thus,

$$\Delta E_{1,2} = \frac{W_{1,2}}{Q_t} \quad (2-15)$$

Here $W_{1,2}$ is the algebraic value of the work done by the external agent in moving the positive test charge Q_t from 1 to 2. For example, if point 2 is at a higher potential than point 1, the external agent will do positive work in pushing the test charge up to point 2.

A useful application of the potential increase will be apparent if we consider the work done while an ion moves between two points 1 and 2 in an electric field. From Eq. (2-15), we find that the work done by the external agent while the ion of charge

$+q$ moves from 1 to 2 is equal to

$$W_{1,2} = +q \Delta E_{1,2} \quad (2-16)$$

Usually an ion moves under the influence of the forces of the field, and it is desirable to calculate the work done *by these electrostatic forces*. This work is equal but opposite in sign to that given by Eq. (2-16).

2-4. Relation between Potential and Electric Intensity.—

Since electric intensity is defined in terms of the force on a test charge and the potential is defined in terms of the work done on a test charge, there will be a relation between intensity and potential that is based on the relation of force to work. Figure 2-5 shows the conditions in an electric field when a test charge $+Q_t$ is moved a distance dl from B to C along the path $ABCD$. Let F represent the intensity at point B , f represent the electro-

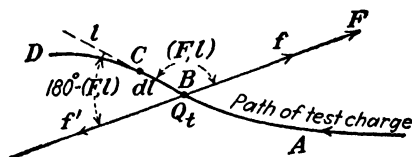


FIG. 2-5.

static force on the test charge, and f' the equal and opposite external force. According to Eq. (2-15) the potential increase dE from B to C is given by the ratio of the work done by the external force during the displacement dl divided by the test charge; i.e.,

$$dE = \frac{W_{B,C}}{Q_t} = \frac{f' \cos [180 - (F, l)] dl}{Q_t} \quad (2-17)$$

Here (F, l) is the angle between the direction of the intensity F and the positive direction of dl , indicated by the line Bl in Fig. 2-5. We note that f'/Q_t has the same magnitude as f/Q_t and therefore also equals the magnitude of the intensity F . Consequently, Eq. (2-17) reduces to the following:

$$dE = F \cos [180 - (F, l)] dl$$

It is more convenient to express this relation in terms of the cosine of (F, l) , thus:

$$dE = -F \cos (F, l) dl \quad (2-18)$$

Upon integrating Eq. (2-18), we obtain a formula for computing the potential increase from a point A to any other point B on the path of the test charge

$$\Delta E_{A,B} = - \int_A^B F \cos (F,l) dl \quad (2-19)$$

This integral is useful for finding the potential difference in a situation where the intensity variation is known. Such an application of this relation will be made in Sec. 2-5 in the concentric-cylinder problem.

Since intensities are derived from known potential distributions more often than the reverse, it will be helpful to obtain from Eq. (2-18) expressions that may be used for this purpose. Thus, the *component* of the electric intensity in the direction of positive l is

$$F_l = F \cos (F,l) = - \frac{dE}{dl} \quad (2-20)$$

If the path lies in the direction of the intensity, the potential is decreasing and, furthermore, is decreasing at a maximum rate compared with that along any other path through the given point. Thus, it is possible to find the intensity at a point by finding the maximum rate of decrease of potential; *i.e.*,

$$F = \left(- \frac{dE}{dl} \right)_{\max} \quad (2-21)$$

As a simple illustration of these relations, consider the situation shown in Fig. 2-1 where a single charge of $+10^{-8}$ coulomb is setting up a field. The potential at points in the region outside of the sphere is given¹ by formula (2-11), $E = Q/4\pi\epsilon r$. By inspection we see that the maximum rate of decrease of potential is in the radial direction; therefore, in accordance with Eq. (2-21) we write

$$F = - \frac{dE}{dl} = - \frac{dE}{dr} = \frac{Q}{4\pi\epsilon r^2}$$

This is, of course, the same value as that previously obtained for the concentrated charge [Eq. (2-6)]. If we assume that the

¹ For points outside a uniformly charged sphere the charge acts as if it were concentrated at the center of the sphere. See general reference 2-2, p. 132.

sphere in Fig. 2-1 has a radius of 0.02 m., the potential and intensities on a radial line have values as shown in Fig. 2-6.

In a three-dimensional rectangular system of coordinates the potential E would be expressed as a function of the three coordinates x , y , and z . If we let F_x represent the component of the intensity in the x -direction, we see from Eq. (2-20) that

$$F_x = -\frac{\partial E}{\partial x} \quad (2-22)$$

Similar expressions hold for the components F_y and F_z . Thus, if the potential distribution is known in a region, the component

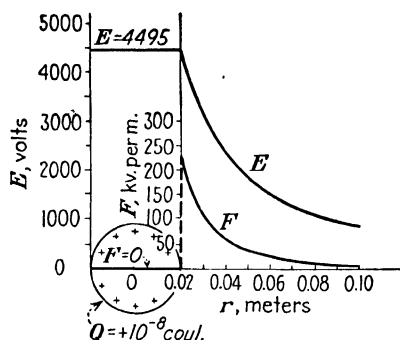


FIG. 2-6.—Potential and electric intensity variations in the neighborhood of a charged sphere.

intensities can be calculated, and the resultant intensity can be obtained by vector addition.

As an example of this method, we propose to calculate the y -component of the electric intensity at point P in Fig. 2-3. The potential at points on the y -axis is given by Eq. (2-13). To find F_y , we calculate the negative of the partial derivative of (2-13) with respect to y , thus:

$$\begin{aligned} F_y &= -\frac{\partial E}{\partial y} = -\frac{q}{4\pi\epsilon} \frac{d \left[\log \left(\frac{s + \sqrt{s^2 + y^2}}{y} \right) \right]}{dy} \\ &= -\frac{q}{4\pi\epsilon} \frac{1}{\frac{s + \sqrt{s^2 + y^2}}{y}} \frac{d \left(\frac{s + \sqrt{s^2 + y^2}}{y} \right)}{dy} = \frac{q}{4\pi\epsilon} \frac{s}{y \sqrt{s^2 + y^2}} \quad (2-23) \end{aligned}$$

This result agrees with the value previously obtained [Eq. (2-7)].

2-5. Gauss's Theorem.—Let us calculate the surface integral of the electric intensity F over the closed surface S taken around a

concentrated charge Q as shown in Fig. 2-7a. The operation of taking the surface integral consists in dividing the surface into elementary areas dA , multiplying these areas by the normal component F_n of the electric intensity at each of the areas, and summing these products over the whole surface. In Fig. 2-7a the direction of the normal to the surface at dA is indicated by the line marked n . Since the normal component F_n might be either outward or inward through the surface, we specify an

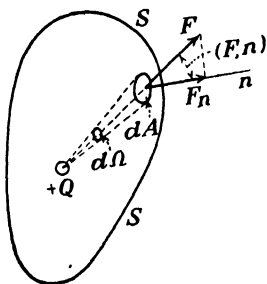


FIG. 2-7a.

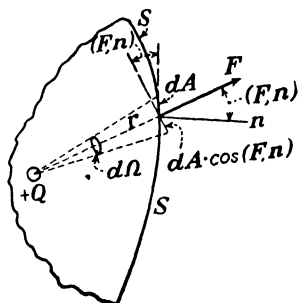


FIG. 2-7b.

outward normal component as positive and an inward component as negative. Thus, the mathematical form of the surface integral is the following:

$$\text{Surface integral of } F = \int_{\text{closed surface}} F_n dA = \int_{\text{closed surface}} F \cos (F, n) dA \quad (2-24)$$

Here the angle (F, n) is the angle between the positive direction of F and the positive (outward) normal to the surface.¹ Upon replacing F by its value given in Eq. (2-6) we obtain

$$\text{Surface integral of } F = \frac{Q}{4\pi\epsilon} \int_{\text{closed surface}} \frac{\cos (F, n) dA}{r^2} \quad (2-25)$$

To evaluate the integral given in (2-25), consider the *solid* or *conical* angle formed by drawing lines from the boundary of the surface element dA to the charge Q . The cone thus formed bounds the solid angle $d\Omega$, whose magnitude is, by the definition

¹ We note that though F_n is algebraic, *i.e.*, carries its own sign, the symbol F represents a positive number, the magnitude of the electric-intensity vector. Thus the sign of $\cos (F, n)$ automatically satisfies the convention regarding the positive normal component of F .

of a solid angle, equal to the ratio of the area $d\sigma$ cut out by the cone from a sphere of radius r to the square of the radius;¹ i.e.,

$$d\Omega = \frac{d\sigma}{r^2} \quad (2-26)$$

The angle $d\Omega$ is measured in steradians. To evaluate the elementary solid angle $d\Omega$ in Fig. 2-7a, consider Fig. 2-7b, which is a view from such a direction that Q , F , and n are in the plane of the paper. Since the projection of the area dA on the sphere of radius r is $dA \cos(F, n)$, we see that $d\Omega$ has the value

$$d\Omega = \frac{dA \cos(F, n)}{r^2} \quad (2-27)$$

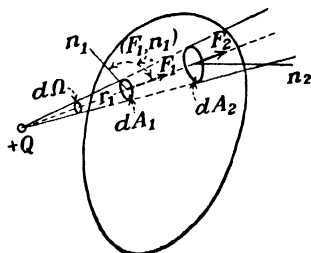


FIG. 2-8.

In view of this relation, the surface integral (2-25) may now be written

$$\text{Surface integral of } F = \frac{Q}{4\pi\epsilon} \int_0^{4\pi} d\Omega = \frac{Q}{\epsilon} \quad (2-28)$$

This states that the surface integral of F is equal to the quantity of charge contained within the surface, divided by the permittivity of the medium.

Before generalizing this relation, we examine the effect of a charge located outside the closed surface on the value of the surface integral of the intensity. Thus, in Fig. 2-8, we consider an elementary area dA_1 and draw a cone from the boundary of dA_1 to the charge Q . If we extend this cone, it cuts out the area dA_2 as it emerges from the closed surface. The contribution to the surface integral for these two elementary areas will be

$$F_1 \cos(F_1, n_1) dA_1 + F_2 \cos(F_2, n_2) dA_2 \quad (2-29)$$

¹ The surface area σ cut out by a given cone from any sphere centered at the vertex of the cone is proportional to the square of the radius of the sphere. If the radius of the sphere is taken as unity, the solid angle is identically equal to the area cut out of the unit sphere [see defining equation (2-26)]. Since the total surface area of the unit sphere is 4π , the total solid angle about a point is 4π steradians.

The values of F_1 and F_2 may be inserted in expression (2-29) to obtain

$$\frac{Q}{4\pi\epsilon} \frac{\cos (F_1, n_1) dA_1}{r_1^2} + \frac{Q}{4\pi\epsilon} \frac{\cos (F_2, n_2) dA_2}{r_2^2} \quad (2-30)$$

We again note that $dA_1 \cos (F_1, n_1)$ is the projected area of dA_1 on a sphere of radius r_1 and consequently

$$\frac{\cos (F_1, n_1) dA_1}{r_1^2} = -d\Omega$$

Here $d\Omega$ is the solid angle of the cone, and the minus sign appears because the angle (F_1, n_1) is an obtuse angle. Similar reasoning shows that

$$\frac{\cos (F_2, n_2) dA_2}{r_2^2} = d\Omega$$

Consequently, expression (2-30) reduces to

$$-\frac{Q}{4\pi\epsilon} d\Omega + \frac{Q}{4\pi\epsilon} d\Omega = 0 \quad (2-31)$$

That is, the contributions to the surface integral over the areas dA_1 and dA_2 exactly cancel each other. Since the whole surface can be accounted for by cones similar to the one shown in Fig. 2-8, it follows that the surface integral of the intensities caused by a charge outside the closed surface is zero.

When a number of charges $Q_1, Q_2, Q_3, \dots$ are contained within a closed surface, the surface integral of the electric intensity will consist of the sum of terms of the type given in Eq. (2-28) for each of the charges. This law may be written as follows:

$$\int_{\text{closed surface}} F \cos (F, n) dA = \frac{Q_1 + Q_2 + Q_3 + \dots}{\epsilon} = \frac{\Sigma Q}{\epsilon} \quad (2-32)$$

This is easily proved if we regard the total normal component of the intensity as the sum of the normal components caused by each of the charges. Therefore, the total surface integral can be thought of as the sum of the surface integrals from each of the components taken separately, and each of these integrals contributes a term as given in Eq. (2-28).

Equation (2-32) is a mathematical form of Gauss's theorem. It states that *the surface integral of the electric intensity over a*

closed surface is equal to the net charge contained within that surface divided by the permittivity of the medium. A slight extension of the foregoing proof of this theorem will show that the charges within the surface may be distributed on conductors or as space charges as well as in the form of concentrated charges. The symbol ΣQ in Eq. (2-32), therefore, is considered to mean the algebraic value of the net charge within the surface.

Gauss's theorem is a useful tool for finding the electric intensity in certain types of problem. The method employed and the types of problem that can be solved are illustrated by the following examples.

The first problem is that of finding the intensity variation in a spherical region which contains a uniform space charge. Let the space charge have a density of $+\rho$ coulombs per unit volume throughout a sphere of radius R (Fig. 2-9). To apply Gauss's theorem, we imagine a spherical surface of radius r . From the symmetry of the geometry, we see that the intensity vector F is constant in magnitude over this sphere and is radially directed. Therefore, the surface integral of F reduces to the constant F multiplied by the surface area of the sphere, or $F \times 4\pi r^2$. The net charge contained within the sphere of radius r is the volume density of the charge multiplied by the volume of the sphere or $\rho \frac{4}{3}\pi r^3$. Applying Gauss's theorem in the form of Eq. (2-32), we find

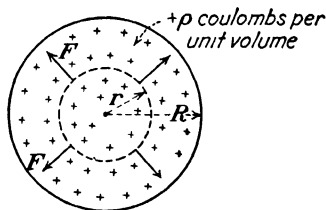


FIG. 2-9.

$$F4\pi r^2 = \frac{\rho \frac{4}{3}\pi r^3}{\epsilon}$$

This reduces to

$$F = \frac{\rho r}{3\epsilon} \quad (2-33)$$

Thus, the intensity at points within the spherical space charge varies linearly with the radius r .

A repetition of the procedure used above, with an imaginary sphere of radius r greater than the radius of the spherical space charge, will yield the following value for the intensity at points

outside the space charge:

$$F = \frac{\rho R^3}{3\epsilon r^2} \quad (2-34)$$

This relation is, of course, the same as that given by Eq. (2-6) if Q is replaced by the total quantity of space charge. Thus, for points outside the spherical charge, the intensity is the same as if all of the charge were concentrated at the center of the sphere.

As a second example, consider the problem of finding the intensity and potential variations between two concentric conducting cylinders of infinite length (Fig. 2-10). Let the inner cylinder of radius R_1 be charged with q coulombs per unit length and the outer cylinder of radius R_2 with $-q$ coulombs per unit length. We choose the imaginary cylinder of radius r and of unit length shown in Fig. 2-10 for the closed surface over which to apply Gauss's theorem. From the symmetry of the problem, it

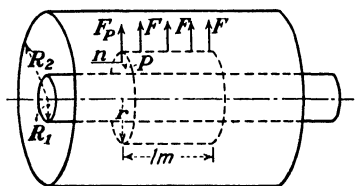


Fig. 2-10.—Field between coaxial cylinders.

is seen that the intensity vectors will be radially directed. Thus, at point P , on an end of the cylinder, the intensity is perpendicular to the normal n , and the normal component is zero. Furthermore, the intensity inside the inner conducting cylinder is zero. Therefore, the

contribution to the surface integral from the ends of the cylinder is zero. The intensity vectors over the outer surface of the imaginary cylinder are normal to the surface and of constant value. Let this value be F ; then, since the area of the outer surface is $2\pi r$, the surface integral reduces to $F \times 2\pi r$. The quantity of charge enclosed by the closed surface is q , and so Eq. (2-32) reduces to

$$F \times 2\pi r = \frac{q}{\epsilon}$$

Therefore,

$$F = \frac{q}{2\pi\epsilon r} \quad (2-35)$$

Since the potential difference, rather than the charge, is usually the known quantity, we apply Eq. (2-19) to find the potential increase $\Delta E_{1,2}$ from the inner to the outer cylinder. Since F is

radially directed, Eq. (2-19) reduces to the following:

$$\Delta E_{1,2} = - \int_{R_1}^{R_2} F \, dr$$

Introducing the known value of F given in Eq. (2-35),

$$\Delta E_{1,2} = - \int_{R_1}^{R_2} \frac{q}{2\pi\epsilon} \frac{dr}{r} = - \frac{q}{2\pi\epsilon} \log \frac{R_2}{R_1}$$

Since we are interested only in the potential difference E between the electrodes, we disregard the minus sign and then have

$$E = \frac{q}{2\pi\epsilon} \log \frac{R_2}{R_1} \quad (2-36)$$

If we eliminate q between Eqs. (2-35) and (2-36), we have

$$F = \frac{E}{r \log (R_2/R_1)} \quad (F \text{ is radially directed}) \quad (2-37)$$

This relation shows that the intensity drops off inversely as the radius increases; and since the intensity inside the inner cylinder

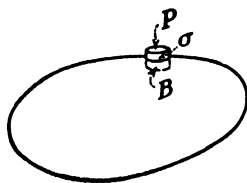


FIG. 2-11a.

Intensity just outside of a conducting surface.

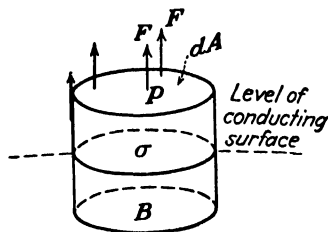


FIG. 2-11b.

is zero, there is a maximum value of intensity just outside the inner cylinder.

The electric intensity just outside any conducting surface that has a known charge density can readily be deduced by the use of Gauss's theorem. Consider a point P that is very close to a conducting body as shown in Fig. 2-11a. The charge density just below P has the known value of σ coulombs per unit area. As a surface over which to apply Gauss's theorem, we choose a cylindrical box with its top including point P and its bottom below the surface of the conductor. An enlarged view of this box is shown in Fig. 2-11b. If point P is only an infinitesimal

distance from the conducting surface, the intensity vectors on the top surface will be constant and normal to the surface. The intensity vectors at the sides of the box will be tangent to the surface, or zero below the conducting surface. The only contribution to the surface integral of the intensity is found over the area of the top of the box dA . The net charge enclosed is σdA . Consequently,

$$F dA = \frac{\sigma dA}{\epsilon}$$

or

$$F = \frac{\sigma}{\epsilon} \quad (2-38)$$

Thus, the intensity just outside a conducting surface is directly proportional to the charge density on the surface.

As a concluding example, consider the electric field between two parallel conducting plates of infinite extent, one of which is

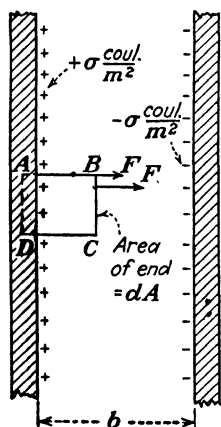


FIG. 2-12.—Field between parallel plates.

charged with $-\sigma$ coulombs per unit area and the other with $+\sigma$ coulombs per unit area (Fig. 2-12). We can apply Eq. (2-38) and state that the intensity between the plates and very close to either one is normal to the plates and has the value σ/ϵ . To show that the intensity between the plates has this same value at all points, we apply Gauss's theorem over the surface of a right cylinder of any length x that has its elements normal to the plates and has one end buried in a positively charged plate (see $ABCD$ in Fig. 2-12). Considerations of symmetry dictate that the intensity will be normal to the plates and constant over the outer end BC of the cylinder. Since the intensity must be zero in the conductor, we see that the only

contribution to the surface integral is over the outer end, whose area is dA . The charge enclosed by the cylinder is σdA . Therefore, we can apply Gauss's theorem to find that $F = \sigma/\epsilon$, and F is constant everywhere between the plates.

The potential difference between the plates can be calculated in terms of the known separation b between the plates. Thus, by

applying Eq. (2-19) we find

$$E = \frac{\sigma b}{\epsilon}$$

The electric intensity between the plates can therefore be expressed as follows:

$$F = \frac{E}{b} \quad (2-39)$$

This relation also holds reasonably well in the space between *finite* plates at points that are not too near the edges, say at points not closer than twice the separation. Unless special precautions are taken,¹ charges will concentrate at the edges of the plates and thereby increase the intensity at the edge above the value found elsewhere on the electrode.

2-6. Mapping Electric Fields.—The methods discussed in the foregoing sections enable us to calculate the electric intensity and potential distribution in the simpler fields. In this section, we shall discuss additional methods used in visualizing electric fields and shall refer to some methods for determining the field distribution.

As already mentioned in Sec. 2-2, the directions of the intensity vectors may be indicated by lines of electric intensity which can be drawn in a picture of an electric field. The value of such a picture is enhanced if an indication of the potential variation in the field is also given by showing the location of several equipotential surfaces in the field. As its name implies, an *equipotential surface is a surface at all points of which the potential is the same*. The field picture, or map, is usually shown in a cross-sectional view, so that the equipotential surfaces are seen as lines. For example, see Figs. 2-13a and 2-13b, which show the field maps that have been calculated for two small charged bodies of unlike sign (Fig. 2-13a) and of like sign (Fig. 2-13b).

Because of the way in which lines of electric intensity and equipotential surfaces have been defined, it follows that *the lines of electric intensity cross the equipotential surfaces at right angles*. For if the electric intensity were not normal to an equipotential surface, the force acting on a test charge placed on the surface would have a component parallel to the equipotential surface. Since no work is done in moving a test charge

¹ See ROGOWSKI, W., *Arch. Elektrotech.*, **12**, 1-15, 1923.

about on an equipotential surface, the component of force parallel to the surface must be zero, and the intensity must be normal to the surface.

Furthermore, the *lines of electric intensity* enter the surface of conductors at right angles, since the surface of a conductor is an equipotential surface.

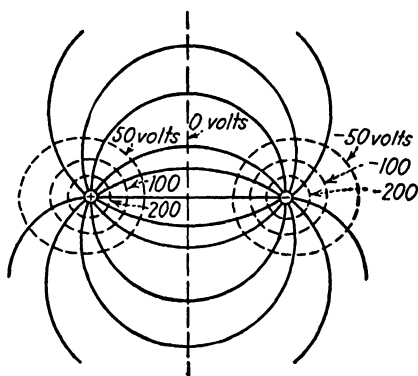


FIG. 2-13a.—Field of equal and opposite charges, each of 10^{-9} coulomb.

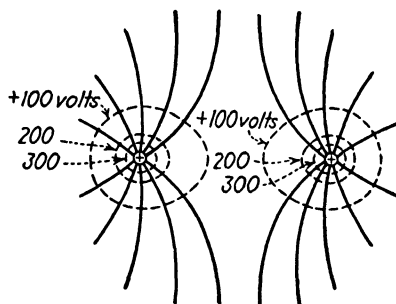


FIG. 2-13b.—Field of two charges, each of 10^{-9} coulomb.

Having briefly discussed the field map and some of its properties, we may reexamine the general methods of determining the properties of electric fields. The problem that presents itself is often the following: Given a configuration of electrodes each of which has a known potential, we may wish to find the electric intensities and potentials in the region around the electrodes. Then, again, the initial information may consist of known charge distributions on the given set of electrodes. The

methods of solving such problems may be classified under the following heads:

1. Mathematical analysis.
2. Graphical mapping.
3. Experimental reproduction.

In general, the application of analytical methods is limited to problems in which the electrodes have simple geometrical shapes. In practical situations the mathematical treatment often becomes difficult if not impossible. Special mathematical methods, such as the method of conformal representation,¹ extend the range of problems that can be treated. Another special method, Kelvin's *method of images* which is applicable in certain cases, is discussed in the next section.

The graphical method² of field mapping finds much use in the solution of complicated electric or magnetic fields such as are encountered in the design of electrical machinery and of high-voltage insulators. In the case of the electric fields, the method consists in first sketching the lines of electric intensity and equipotential surfaces in accordance with the general laws that these obey, with the help afforded by previous experience with portions of the field under consideration or with similar fields. This trial map of the field can then be refined by a repetition of the same procedure until an accurate map is obtained. This method has been used mainly in two-dimensional fields, *i.e.*, in fields in which the electrodes consist of exceedingly long cylinders.

The experimental method of determining the field map consists in reproducing either the original field or an analogous field in an experimental model in which the equipotential surfaces can be ascertained by test. The procedure most often employed is to use the conducting field as the analogue of the electrostatic field. In this method a conducting electrolyte, usually water, fills the space between metal electrodes. The conductivity of the metal relative to that of the water is about 10^8 , so that the electrodes may be taken to be equipotential surfaces. To reproduce a two-

¹ SOKOLNIKOFF, I. S., and E. S. SOKOLNIKOFF, "Higher Mathematics for Engineers and Physicists," Chap. XV, McGraw-Hill Book Company, Inc., New York, 1934. Dow, W. G., "Fundamentals of Engineering Electronics," Chap. II, John Wiley & Sons, Inc., New York, 1937. WEBER, E., *Elec. Eng.*, **53**, 1563-1570, 1934.

² General reference 2-1; in particular, Chap. VII.

dimensional problem a shallow insulating tray may be used to hold the water; but if a three-dimensional problem is to be reproduced, a tank of relatively large dimensions must be provided. Figure 2-14 shows a reproduction of a two-dimensional field, with cylindrical electrodes B and C mounted in a shallow tray of uniform depth. The alternator A , preferably having a frequency of 400 to 800 cycles/sec., is connected to the two electrodes. By using an audible frequency, it is possible to use telephone receivers as the detector and to eliminate errors that might be caused by polarization voltages at the electrodes.

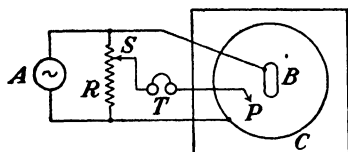


FIG. 2-14.—Electrolytic method of field mapping.

The potentiometer R divides the voltage, so that the probe P is raised to some potential between that of C and B determined by the setting of the slider S . When the probe is dipped in the water, a sound will be heard in the receivers T , unless the probe hap-

pens to be on a surface at the same potential as the slider S . By moving the probe about in the water, meanwhile keeping the sound at a minimum, an equipotential surface can easily be traced out.

After a number of the equipotential surfaces have been mapped, the field-intensity lines can be sketched by using the rules that have been laid down. If a sufficient number of equipotential surfaces have been determined, a highly accurate field map can be drawn.¹ After the field map has been completed, the value of the intensity at any point on a line of intensity can be found by calculating the rate of change of potential with distance *measured along the line; i.e.,*

$$F = - \frac{dE}{dl} \quad (2-21)$$

The value of dE/dl in Eq. (2-21) is most readily obtained from an auxiliary Cartesian graph of E as a function of l .

2-7. The Method of Images.²—This method, originated by Lord Kelvin in 1848, is applicable to a few special problems in

¹ For examples of field maps determined in this way, see Figs. 3-6a and 3-13.

² General reference 2-1, Chap. V; general reference 2-2, Sec. 112.

which a given distribution of known charges sets up an unknown system of *induced* charges on the conducting bodies in the field. Then an auxiliary charge or system of charges is sought which, acting together with the original known charges (but without the actual induced charges), will reduce the potentials at the boundaries of the original field, *i.e.*, at the surfaces of the conductors, to the correct values. If such a system of charges can be found and if the solution is known for the field of the original charges plus the auxiliary charges, then the problem is solved. This follows because an electrostatic field is uniquely determined by specifying the potentials of the conductors in the field and the free charges in the field. Any solution that reproduces the correct potentials of the conductors in the presence of the free charges is, therefore, the unique solution of the problem. The

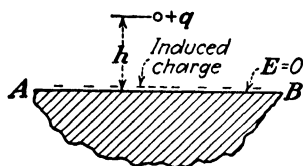


FIG. 2-15a.

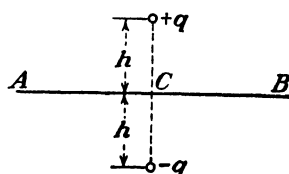


FIG. 2-15b.

following example, which occurs in the study of the emission of electrons from metals, will serve as an introduction to the method.

Consider a concentrated charge of $+q$ located as shown in Fig. 2-15a a distance h from the plane surface of an extended conducting body whose potential is zero. The conducting body could be replaced by a thin plate without changing the solution of the problem. We wish to find the electric intensity and potential variation in the field; but since the distribution of induced charge is as yet unknown, the problem cannot be solved by the usual methods. If we recall Fig. 2-13a, we see at once that, if an equal and opposite charge $-q$ is placed a distance h below the surface AB , then this auxiliary charge acting with the original charge $+q$ (but without the actual induced charge) will reduce the surface of the conductor to the correct potential $E = 0$. Consequently, the system of charges $+q$ and $-q$ in Fig. 2-15b is the equivalent, as far as the field above plane AB is concerned, of the charge $+q$ with its induced charge on the metal surface. The charge $-q$ is called the electrical image of the charge $+q$.

The induced charge density on the surface of the conductor can now be evaluated. This is done by using the inverse-square law to calculate the intensity F at the surface of the plane AB set up by the two charges $+q$ and $-q$. Then, the relation in Eq. (2-38) can be used in the form

$$\sigma = \epsilon F$$

to obtain the surface density of charge that exists at the corresponding point on the surface of the conductor.

The force of attraction between the charge $+q$ and the induced charge on the plane AB can also be calculated by means of the equivalent system of charges in Fig. 2-15b. Thus, this force of attraction is given by Coulomb's law, as

$$f = \frac{q^2}{16\pi\epsilon h^2} \quad (2-40)$$

Another way to obtain this expression is to calculate the electric intensity F at the point occupied by $+q$ and caused by $-q$ and then apply the relation $f = Fq$.

Problems

1. An isolated conducting sphere of 0.02 m. radius is charged to a potential of 1,000 volts. (a) Describe the equipotential surfaces in the field for which $E = 1,000$, $E = 700$, $E = 400$, and $E = 100$ volts. (b) Draw a curve showing the variation of electric intensity with distance from the surface of the sphere to a point 0.2 m. from the center of the sphere. Draw a curve showing the potential variation over the same range.

2. Find the greatest charge that can be carried by a metal sphere of 0.1 m. radius if the sphere is placed in air remote from other objects. Assume that the dielectric strength of air is 2,000 kv./m.

3. A parallel-plate air condenser has an air gap of 1 mm. If the potential difference between the plates is 750 volts, what is the electric intensity in the gap? What force would be exerted on an electron placed in this air gap?

4. Two small, charged, conducting spheres A and B are arranged and charged as shown in Fig. 2-16. (a) Plot curves showing the variation of electric intensity and potential along the x -axis, and also the same curves for the y -axis. Take in the range from $x = -0.2$ to $x = +0.2$, and $y = -0.2$ to $y = 0.2$. (b) Explain the relation that exists between the intensity and potential curves for the x -axis. Does the same relation hold for the curves for the y -axis?

5. A concentrated charge Q is located at the center of a spherical surface S . If the value of the integral $\int F \cos (F, n) dA$ taken over the surface S is 339, what is the magnitude of Q ?

6. The electric-intensity vector $\mathbf{F}$ is constant in magnitude and direction in a given region. If the value of the integral $\int \mathbf{F} \cos(\mathbf{F}, \mathbf{n}) dA$ taken over a plane area of 2 sq. m. is 5,200 and the vector $\mathbf{F}$ makes an angle of 30 deg. with the normal to the plane surface, what is the value of $\mathbf{F}$?

7. Suppose that, in the problem illustrated in Fig. 2-17, electrons are transferred from earth to plate P until plate P has a potential of -500 volts relative to earth. (a) What is the charge density on the two sides of P and on the inner faces of A and B ? (b) Sketch curves showing intensity and potential variations on a line perpendicular to the plates that extends from A to B .

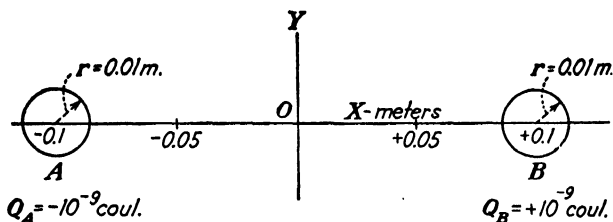


FIG. 2-16.

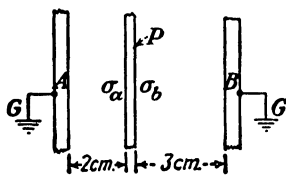


FIG. 2-17.

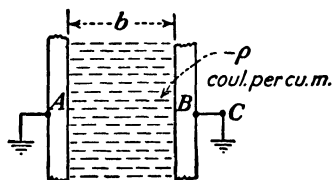


FIG. 2-18.

8. Suppose that a thin, extended, plane metal plate P is mounted (on insulators) in air parallel to and between two grounded metal-plate electrodes A and B as in Fig. 2-17. Suppose that electrons have been transferred from earth G to the plate P until the plate contains a charge of 8.85×10^{-8} coulomb/sq. m. of plate ($\sigma_a + \sigma_b = -8.85 \times 10^{-8}$). Calculate (a) the surface density of charge on the two surfaces of P and on the inner surfaces of A and B , (b) the magnitude and direction of the intensities between the plate P and the electrodes, and (c) the potential of plate P relative to plates A and B .

Sketch curves of electric intensity and potential variation for points on a straight line from A to B perpendicular to the plates.

9. Suppose that a uniform negative space charge of $-\rho$ coulombs per cubic meter is fixed in space between the two extended metal plates A and B in Fig. 2-18. (a) What is the surface density of the induced charge on the plates A and B ? (b) Calculate the variation of electric intensity and potential along a line perpendicular to and between the plates. (c) Sketch curves showing the features of the variation of intensity and potential along this line.

10. Suppose that in the situation shown in Fig. 2-18 the connection is broken at C and a battery is inserted in such a manner as to raise the potential of plate B to $+E_1$ volts. Redetermine the variation of electric intensity and potential between the plates, and sketch curves to show these variations. Assume that the space charge remains fixed as in Prob. 9.

11. Referring to the situation given in Prob. 10, find how large E_1 must be so that if an electron is given off from plate A with very small or infinitesimal velocity it will nevertheless move to plate B . Imagine that the initial space charge remains fixed in space and only the single electron is free to move.

12. Referring to the spherical distribution of space charge shown in Fig. 2-9, find (a) the potential at the center of the sphere and (b) the variation of the potential along a radial line through the sphere.

13. Referring to the concentric-cylinder arrangement shown in Fig. 2-10, find the potential at any point along a radial line between the cylinders. Express the potential as a function of the radius r , the dimensions of the cylinders, and the potential difference ΔE . Consider the outer cylinder to be at zero potential.

14. A very long conducting cylinder of radius R_1 contains a uniformly distributed space charge of $+\rho$ coulombs per cubic meter. What is the nature of the electric intensity and potential variation at points on a radial line from the axis to the cylinder? Sketch curves showing the variation of electric intensity and potential along this line.

15. Calculate the variation of the electric intensity and surface density of charge on an extended, grounded, metal surface caused by a concentrated charge of 4×10^{-9} coulomb placed 0.2 m. above the surface.

16. What is the force of attraction between an electron and its "induced charge" on a metal surface if the electron is 10^{-8} m. distant from the surface? How great would the electric intensity of a field need to be to cause the same force on the electron?

17. Show that the total quantity of induced charge on the infinite metal plate in Fig. 2-15a is equal in magnitude to the original charge $+q$.

General References

- 2-1. ATTWOOD, S. S.: "Electric and Magnetic Fields," 2d ed., John Wiley & Sons, Inc., New York, 1941.
- 2-2. BENNETT, E., and H. M. CROTHERS: "Introductory Electrodynamics for Engineers," McGraw-Hill Book Company, Inc., New York, 1926.
- 2-3. PAGE, L., and N. J. ADAMS, JR.: "Principles of Electricity," D. Van Nostrand Company, Inc., New York, 1931.

CHAPTER III

MOVEMENT OF CHARGED PARTICLES IN HIGH VACUUM¹

3-1. Forces in the Electric Field.—We have defined the electric intensity at a point in an electric field as the ratio of the force to the charge carried by a very small test body placed at the point. From this definition, it follows that if a very small test charge is placed in an electrostatic field the force on the test charge is given by

$$\mathbf{f} = q\mathbf{F}$$

The charge carried by the test particle, q , may be either positive or negative. Stating this another way, we can say that, if a small charged body is introduced into an electric field and if the charge carried by the body is small enough so that it does not alter the original electric field distribution, then the force acting on the charged body is given by the equation above. The particles with which we shall usually be concerned are either electrons or molecular ions. We shall assume for the purposes of this study that these particles are sufficiently small so that the presence or absence of any one particle will not disturb the electric field due to all other charges in the neighborhood.

We have also shown in Sec. 2-4 that the x -, y -, or z -components of the electric intensity at any point may be calculated from the corresponding rate of change of potential in the x -, y -, or z -direction at that point. Thus,

$$\left. \begin{aligned} F_x &= -\frac{\partial E}{\partial x} \\ F_y &= -\frac{\partial E}{\partial y} \\ F_z &= -\frac{\partial E}{\partial z} \end{aligned} \right\} \quad (2-22)$$

¹ We shall assume that the particles are moving in a high enough vacuum so that collisions with molecules of the gas do not occur.

General references for Chap. III will be found on p. 79.

Another means of representing the force or the components of the force on a small charged body placed in an electric field is thus in terms of the potential variation at the point as shown in the following equations:

$$\left. \begin{aligned} f_x &= q \left(-\frac{\partial E}{\partial x} \right) \\ f_x &= -q \frac{\partial E}{\partial x} \\ f_y &= -q \frac{\partial E}{\partial y} \\ f_z &= -q \frac{\partial E}{\partial z} \end{aligned} \right\} \quad (3-1)$$

We see then that the force acting on a small charged particle can be represented in terms either of the electric intensity or of the rate of change of potential.

If we release an electron or other small particle carrying a charge q at some point in an electric field, the force that then acts on the particle is proportional to the charge on the particle and the magnitude of the electric-field intensity. The particle will consequently start to move in the direction of the force and will acquire an acceleration given by the customary Newtonian equation

$$\mathbf{f} = m\mathbf{a}$$

or

$$\mathbf{f} = q\mathbf{F} = m\mathbf{a} \quad (3-2)$$

3-2. Laws of Motion in an Electrostatic Field.—From the preceding expression for the force acting on the charged particle, we can calculate the acceleration of the particle, as follows:

$$\mathbf{a} = \frac{q\mathbf{F}}{m} = \frac{d\mathbf{V}}{dt} \quad (3-3)$$

The acceleration is thus a vector quantity having the same direction as $\mathbf{F}$ if q is positive or opposite to $\mathbf{F}$ if q is negative.

If $\mathbf{F}$ is constant, the acceleration is constant.¹ This means that, if the particle starts from rest, the velocity of the particle at any time t is given by

$$\mathbf{V} = \mathbf{a}t = \frac{q\mathbf{F}}{m} t \quad (3-4)$$

¹ As in practically all calculations involving the motions of charged particles in electric fields, the acceleration of gravity has been neglected.

and the distance traveled by the particle, starting from rest, in the time t is

$$s = \frac{at^2}{2} = \frac{qEt^2}{2m} \quad (3-5)$$

Because of the analogy between Eqs. (3-3), (3-4), and (3-5) and the corresponding equations for the case of a mass M freely falling in a gravitational field, it is quite common to speak of the "free fall" of a charged particle in an electrostatic field or of a charged particle "falling through" an electrostatic field.

It is still more common to speak of a charged particle "falling through" a difference of potential of E volts in considering the motion of a particle in an electrostatic field. For example, if the electric field is due to two electrodes A and B maintained at a difference of potential of E volts, an electron in moving from one electrode A to the other B under the influence of the electric

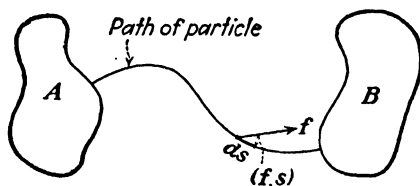


FIG. 3-1.

field (see Fig. 3-1) is often said to have fallen through a difference of potential of E volts.

Let us calculate the energy gained by the particle as it moves through the electric field. If we multiply the force on the charged particle, as given by Eq. (3-2), by the cosine of the angle between f and the direction of the path to get the component of f along the path and then multiply by the incremental length of path ds and take the line integral between the two electrodes A and B , the result, the work done on the particle, is as follows:

$$\begin{aligned} \int_A^B f \cos (f,s) ds &= q \int_A^B F \cos (f,s) ds \\ m \int_A^B a \cos (f,s) ds &= q \int_A^B F \cos (f,s) ds \end{aligned}$$

The acceleration a is in the direction of the force f ; thus $[a \cos (f,s)]$ is the component of the acceleration in the direction of the path and is equal to dV_t/dt , V_t being the component of velocity tangent to the path of the particle. The equation above can

then be rewritten, as follows:

$$\begin{aligned} q \int_A^B F \cos (f,s) ds &= m \int_A^B \frac{dV_t}{dt} ds \\ &= m \int_A^B \frac{ds}{dt} dV_t \\ &= m \int_A^B V_t dV_t \end{aligned}$$

If we now notice that F is in the same direction as f , by definition, and $[F \cos (f,s)]$ is the component of the electric intensity along the path, we can substitute for $[F \cos (f,s)]$ its value as given by Eq. (2-20). In this way, we can easily get

$$-q \int_A^B \frac{dE}{ds} ds = m \int_A^B V_t dV_t$$

We note here that V_t is the same as the total velocity V , for the particle is moving freely under the influence of the applied field. Thus, after carrying out the indicated integration, we obtain

$$qE_{A,B} = \frac{1}{2}mV^2 \quad (3-6)$$

Here $E_{A,B}$ is the difference in potential between A and B , and we see that after the particle "falls through" a difference of potential of E volts its velocity is given by the equation above, [Eq. (3-6)].

Again the analogy with a gravitational field is apparent. At the point A the particle possesses a certain amount of potential energy given by $qE_{A,B}$, analogous to the potential energy possessed by a ball resting on a ledge. A ball with a mass M resting on a ledge at a height h above the earth possesses a potential energy equal to Mgh when the acceleration due to gravity is given by g . If the ball is pushed off the ledge, it is accelerated in the gravitational field. By the time the ball has reached the earth, its potential energy has all been converted to kinetic energy, and in this particular case the relation between kinetic and potential energy is expressed by the following equation:

$$Mgh = \frac{1}{2}MV^2$$

Similarly if a charged particle is held to an electrode A in an electrostatic field, it possesses a potential energy equal to $qE_{A,B}$.

If the charged particle is released, it is accelerated by the electric field; by the time it reaches the point B , all its potential energy has been converted to kinetic energy, and in this case

$$qE_{A,B} = \frac{1}{2}mV^2 \quad (3-6)$$

Let us examine this equation a little more closely. We see that the velocity of the particle after falling through a difference of potential of E volts is entirely independent of the length of the path the particle has traversed and also entirely independent of the form or shape of the electric field. The electric-field intensity may be distorted in any way we please, *i.e.*, may be high at one point, low at another, negative at still another; but still, if the total difference in potential between A and B is $E_{A,B}$ volts, the velocity of the particle at B will be

$$V = \sqrt{\frac{2q}{m} E_{A,B}} \quad (3-7)$$

An electron that has fallen through a potential difference of 1 volt will possess a kinetic energy of 1.6×10^{-19} joule.

$$qE = 1.60 \times 10^{-19} \times 1 = 1.60 \times 10^{-19} \text{ joule} \quad (3-8)$$

This is a convenient unit of energy and is ordinarily called the *electron volt*. Thus, any particle that has an energy of 1.60×10^{-19} joule is said to possess one *electron volt* of energy regardless of how the particle gained its energy. Thus, a ball having a mass of 1 mg. and moving with a velocity of 1 mm./sec. would have a kinetic energy of approximately 3×10^6 e.v. This means that the ball has the same kinetic energy as an electron would have if it fell through a difference of potential of 3 million volts. It does *not* imply that the ball is moving in an electric field or that it is charged.

An electron with a kinetic energy of one electron volt has a velocity of about 5.93×10^5 m./sec.¹ as calculated from Eq. (3-7):

¹ In some cases, writers have found it convenient to express the *velocity* of an electron as well as the *energy* of the electron in electron volts. In these cases the velocity is equal to the velocity that an electron would have after it has fallen through a difference of potential of E volts. In other words, they say that any electron with a velocity of $5.93 \times 10^5 \sqrt{E}$ m./sec. has a velocity of E e.v. and an energy of E e.v. It is obvious that this is a very confusing practice and one that should be avoided.

$$V = \sqrt{\frac{2 \times 1.6 \times 10^{-19}}{9.1 \times 10^{-31}}} \sqrt{E}$$

$$= 5.93 \times 10^5 \sqrt{E} \text{ m./sec.}$$

A hydrogen molecular ion, on the other hand, having a kinetic energy of one electron volt would have a velocity of about 9.85×10^3 m./sec. The calculation is as follows:

$$V = \sqrt{\frac{2 \times 1.6 \times 10^{-19}}{3.3 \times 10^{-27}}} \sqrt{E}$$

$$= 9.85 \times 10^3 \sqrt{E} \text{ m./sec.}$$

If the charged particle was not introduced into the electric field with zero velocity but was projected into the field with an initial velocity V_0 making an angle α with the direction of the electric field as shown in Fig. 3-2, the velocity components at any time t would be given by

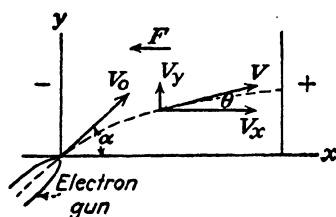


FIG. 3-2.

$$\left. \begin{aligned} V_x &= V_{0x} + \frac{qF}{m} t \\ V_y &= V_{0y} \end{aligned} \right\} \quad (3-9)$$

and the magnitude of the resultant velocity V would, of course, be given by

$$V = (V_x^2 + V_y^2)^{1/2} \quad (3-10)$$

its direction being given by

$$\tan \theta = \frac{V_y}{V_x} \quad (3-11)$$

The distance traveled in the x - and y -directions in time t would be given by

$$\left. \begin{aligned} x &= V_{0x}t + \frac{qF}{2m} t^2 \\ y &= V_{0y}t \end{aligned} \right\} \quad (3-12)$$

In both Eqs. (3-9) and (3-12), V_{0x} and V_{0y} are the x - and y -components of the initial velocity.

The kinetic energy of the particle after falling through a difference of potential of E volts, if it entered the field with an

initial velocity V_0 , would be

$$\frac{1}{2}mV^2 = \frac{1}{2}mV_0^2 + qE \quad (3-13)$$

3-3. Laws of Motion in a Uniform Magnetic Field.—Let us consider now the force on a charged particle moving with uniform velocity in a uniform magnetic field. A typical situation is illustrated in Fig. 3-3. In this simple case, we have a positive charge moving downward with a velocity V at a certain instant. If the lines of magnetic flux are pointing out of the paper, as indicated in the figure, the force acting on the charge will be to the left. If the charged particle is an electron, the force will be in the opposite direction. In either case the magnitude of the force will be given by

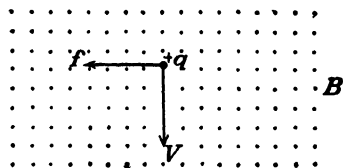


FIG. 3-3.

$$f = BqV \sin (B, V) \quad (3-14)$$

where (B, V) is the angle between the direction of the velocity and the direction of the magnetic flux density vector. In this equation, f is in newtons if B is expressed in webers per square meter, q in coulombs, and V in meters per second.¹ For the case illustrated in Fig. 3-3, the charged particle is moving at right angles to the magnetic field; hence, the force is expressed by the simpler equation

$$f = BqV \quad (3-15)$$

This equation can be derived from the more familiar equation for the force on a conductor carrying current in a magnetic field. That equation for the force is

$$f = BIl$$

if the conductor is at right angles to the magnetic field and takes the form

$$f = BIl \sin (B, l) \quad (3-16)$$

if the conductor makes the angle (B, l) with the magnetic field. Since the current density in any conductor is given by Eq. (1-1)

¹ Equation (3-14) will have the same form in c.g.s. e.m.u. However, f will be expressed in dynes when B is expressed in lines per square centimeter, V is in centimeters per second, and q is in e.m.u.

of Chap. I, we shall have

$$I = NqV \times A$$

Thus

$$\begin{aligned} f &= BNqVal \sin(B,l) \\ &= BqV \times NAl \sin(B,l) \end{aligned}$$

Since NAl is the total number of free electrons in the wire of length l , the force on each electron is

$$f = BqV \sin(B,V) \quad (3-14)$$

and if the electrons are moving at right angles to the field, *i.e.*, if the conductor is at right angles to the field, it is simply

$$f = BqV \quad (3-15)$$

The force determined by any of the preceding equations always acts along a line perpendicular to both the velocity and the direction of the magnetic flux density vector. The direction along this line in which the force acts can be determined by any of the usual rules. One of these can be expressed as follows: The force acting on a positive charge moving in a magnetic field is in the same direction that a right-hand screw would advance if the slot on the head of the screw were turned from a position coinciding with the V vector into one coinciding with the B vector.

Since the force acting on the charged particle is always at right angles to the direction of motion, it follows that, if the initial velocity of the particle is perpendicular to the magnetic flux density, the path traced out by the particle will be an arc of a circle. The radius of curvature of the path can be found from the equation

$$\frac{mV^2}{r} = BqV \quad (3-17)$$

which expresses the equilibrium condition when centripetal force is just equal to the force due to the magnetic field. Thus, the radius of the path is

$$r = \frac{mV}{Bq} \quad (3-18)$$

We readily see that if two charged particles are introduced into a uniform magnetic field with the same velocity but with opposite charge they will trace out circles in opposite directions.

In general, a charged particle enters a magnetic field, not exactly at right angles to the magnetic field, but at some other angle. In this case, only the component of the velocity at right angles to the magnetic field contributes toward turning the particle in a circular path. The component of the velocity parallel to the magnetic field is unchanged. Thus, the path of the particle is a helix as shown in Fig. 3-4. The radius of the

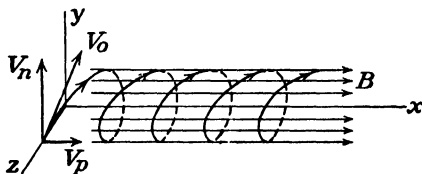


FIG. 3-4.

cylinder on which the particle moves in its helical path is determined only by the component of the velocity normal to the magnetic field. Thus,

$$r = \frac{mV_n}{qB} \quad (3-19)$$

The pitch p of the helix is determined by the time necessary for one cycle (the period) and the velocity component parallel to the field.

$$p = \tau V_p \quad (3-20)$$

or

$$p = \frac{2\pi}{\omega} V_p \quad (3-21)$$

In Eqs. (3-20) and (3-21), τ is the period and ω is the angular velocity.¹ The angular velocity with which the charged particle traverses the circular, or helical, path is quite independent of the direction of the initial velocity. Since the angular velocity in radians per second is given by

$$\omega = \frac{V_n}{r} \quad (3-22)$$

and

$$r = \frac{mV_n}{qB}$$

¹ By the angular velocity ω , we mean the angular velocity of the projection of the moving particle on a plane perpendicular to the axis of the helix [see Eq. (3-22)].

the angular velocity proves to be

$$\omega = \frac{qB}{m} \quad (3-23)$$

The angular velocity thus depends only on the mass of the particle, the charge carried by the particle, and the magnetic-field strength.

As an example of the application of these principles, consider the motion of 400 e.v. electrons traveling at right angles to a magnetic field of 10^{-2} w./m^2 . From Eq. (3-7)

$$V = \sqrt{\frac{2q_e}{m}} \sqrt{E}$$

and, from Eq. (3-18),

$$r = \frac{mV}{q_e B}$$

Therefore,

$$r = \frac{1}{B} \sqrt{\frac{2m}{q_e}} \sqrt{E}$$

or

$$r = 6.7 \times 10^{-3} \text{ m.}$$

The angular velocity is given by Eq. (3-23).

$$\omega = \frac{q_e B}{m}$$

or

$$\omega = 1.76 \times 10^9 \text{ rad./sec.}$$

Thus, the electron would be rotating at the rate of $2.80 \times 10^8 \text{ r.p.s.}$

3-4. The Cathode-ray Oscillograph: the Electrostatic Deflecting Type.—One of the outstanding examples of the application of dynamics of charged particles is the cathode-ray oscillograph. In the cathode-ray oscillograph a beam of electrons is projected usually through an evacuated space to a recording screen, either a fluorescent screen for visual work or a photographic film for direct recording. Between the source of the beam and the screen are placed electrostatic or electromagnetic deflecting electrodes. The beam, then, in passing through the electrodes suffers a deflection or turning which is observed on the screen and which depends directly on the voltage of the electrostatic plates or the current in the magnetic deflecting coils. Although the

beam consists of a stream of electrons, the deflection can be calculated as if a single electron were traveling down the tube. That is, in a high vacuum the electrons make no collisions with gas molecules and hence travel in straight lines unless deflected by an electric or a magnetic field. The only influence the electrons have upon one another is to produce a certain amount of spreading of the beam as it passes down the tube due to the

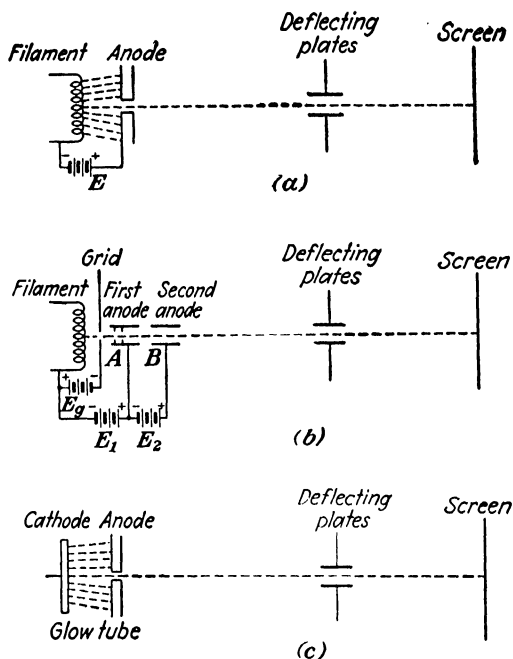


FIG. 3-5.—Types of cathode-ray oscillographs.

electrostatic repulsion of individual electrons. This can be corrected however by suitable focusing devices which will be discussed in Sec. 3-8.

The electron beam may be produced in several ways. One method is to use a thermionic filament as an electron source. In this case the electrons that leave the filament are accelerated toward an anode or a plate by means of a difference of potential as shown in Fig. 3-5a. Some of the electrons pass through a small hole in the anode and continue on in a relatively field-free region until they pass through the deflecting plates and hit the

screen. In other oscillographs the electrons are accelerated twice after leaving the cathode, once in passing from the filament to the first anode and again in passing from the first anode to the second anode. In this arrangement, shown schematically in Fig. 3-5*b*, the two anodes serve as an electron lens and permit the beam to be focused on the fluorescent screen (see Sec. 3-8). A grid is usually incorporated in these tubes, also, so that the intensity of the beam can be varied in very much the same way that the current in an ordinary triode is varied by changing the grid potential. (For a discussion of grid control, see Sec. 6-9.)

With either of these arrangements, it is possible to evacuate the oscillograph tube to a high degree,¹ for all the electrons come initially from a thermionic cathode, and the presence of gas will produce only dispersion of the beam.* This is the type of electron source that is used in most "oscilloscopes." In these tubes the accelerating potential is not more than a few thousand volts, and observations with the oscillograph are made by observing the motion of the cathode-ray beam on a fluorescent screen. With this tube, single transients can sometimes be photographed by focusing a camera on the fluorescent screen, but very fast transients can be observed only if they are repeated periodically on the screen. Single transients that last less than 10 or 20 microsec. cannot be observed with the usual oscilloscope tubes.

When it is necessary to observe transient voltages that persist for less than 10 or 20 microsec., it is necessary to use a higher energy electron beam. Such an oscillograph is usually made so that the photographic film can be introduced directly into the vacuum chamber, and records are made directly by the electron beam as it hits the photographic emulsion. Electron energies in these oscillographs go as high as 60,000 or 80,000 volts, and it is possible to investigate single transients that last only a few hundredths of a microsecond. Transients of this kind are often met in studying lightning disturbances on high-voltage transmission lines. As we shall see a little later and as we already suspect from our previous discussion of electron dynamics, the voltage sensitivity or deflection observed on the screen with a given change in the voltage applied to the deflecting plates of

¹ The importance of small amounts of gas in some tubes for focusing purposes is discussed in Sec. 3-8.

these high-voltage instruments is much less than for the low-voltage oscilloscopes. High-voltage electrons are usually obtained from a cold-cathode glow-discharge tube. Since the voltage drop of a glow discharge increases very rapidly as the pressure is decreased, it is easy to obtain discharge-tube voltages of approximately 60,000 volts. Electrons that leave the cathode of such a tube are thus accelerated and reach the anode with energies of 60,000 e.v. Some of the electrons pass through a very small hole in the anode and pass on through the deflecting plates to the screen or photographic film. This means of acceleration is shown schematically in Fig. 3-5c.

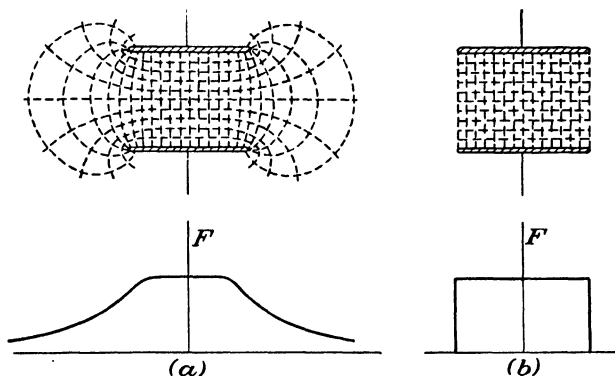


FIG. 3-6.—Electric field distribution between parallel plates: (a) actual; (b) approximate.

Let us consider now the effect on the beam of a potential difference E_d applied to the electrostatic deflecting plates. As the beam enters the field of the plates, each electron in the beam and consequently the beam as a whole suffers an acceleration toward the more positive of the two plates. The actual field between the plates is somewhat as shown in Fig. 3-6a, and a force begins to act on the electrons even while they are still some distance from the plate. The importance of this “fringing” effect at the ends of the electrodes depends upon the relative separation of the plates compared with their length. In many problems, however, the field between the plates can be considered a uniform one as shown in Fig. 3-6b if an equivalent plate length larger than the actual plate length is used.

The force acting on each electron in the beam was found to be

$$f = ma = Fq_e$$

and since the field is a uniform one between plane parallel plates,

$$F = \frac{E_d}{b} \quad (3-24)$$

if b is the separation of the plates. Thus, the force can be written

$$f = \frac{E_d q_e}{b} \quad (3-25)$$

This force begins to act as soon as the electron reaches the end of the deflecting plates, and it continues to act until the electron passes out from between the plates. The total deflection of the electron while between the plates is simply

$$s = \frac{1}{2} at^2 = \frac{q_e E_d}{2mb} t^2 \quad (3-26)$$

The time of transit between the plates depends on the initial velocity of the electrons, and consequently the deflection between

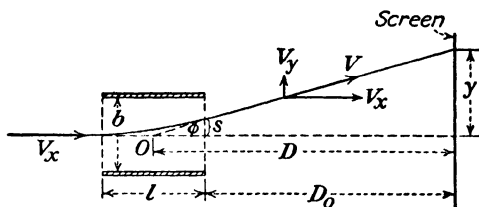


FIG. 3-7.—Electrostatic deflection of cathode-ray beam.

the plates will depend on the energy of the electrons in the beam. Since the electrons are experiencing a force all the time they are between the deflecting plates, they will leave the deflecting plates with a component of velocity perpendicular to their original velocity as shown in Fig. 3-7. The path of the beam between the plates is curved; but outside it is straight, for there the electrons are experiencing no force and the velocity is constant.

With the situation shown in Fig. 3-7 the electrons enter the deflecting field with a velocity V_x given by their initial energy

$$V_x = \sqrt{\frac{2q_e}{m}} \sqrt{E_a} \quad (3-27)$$

This velocity component is not changed. The initial y -component of velocity is zero; but as the electron leaves the plates,

it has a y -component of velocity given by

$$V_y = at = \frac{q_e E_d}{mb} t \quad (3-28)$$

where t is the time of transit between the plates. If the length of the plates is l , the time of transit between the plates is

$$t = \frac{l}{V_x} \quad (3-29)$$

Thus,

$$V_y = \frac{q_e E_d}{mb} \frac{l}{V_x} \quad (3-30)$$

The angle ϕ between the emergent beam and the horizontal is given by

$$\tan \phi = \frac{V_y}{V_x} \quad (3-31)$$

and the deflection that is actually observed on the screen is

$$y = D_0 \tan \phi + s$$

or

$$y = D_0 \frac{V_y}{V_x} + s \quad (3-32)$$

Here, D_0 is the distance from the screen to the end of the deflecting plates, and s is given by Eq. (3-26). Combining Eqs. (3-26) and (3-29) to eliminate t and substituting in Eq. (3-32), we get

$$y = D_0 \frac{V_y}{V_x} + \frac{q_e E_d l^2}{2mb} \frac{1}{V_x^2}$$

If we now substitute for V_x and V_y their corresponding values in terms of E_a and E_d , we obtain

$$\begin{aligned} y &= D_0 \frac{l E_d}{2b E_a} + \frac{l^2}{4b} \frac{E_d}{E_a} \\ &= \frac{l}{2b} \frac{E_d}{E_a} \left(D_0 + \frac{l}{2} \right) \end{aligned} \quad (3-33)$$

But the quantity $\left(D_0 + \frac{l}{2} \right)$ represents the distance from the screen to the center of the deflecting plates. Thus,

$$y = D \frac{l}{2b} \frac{E_d}{E_a} = K_1 \frac{E_d}{E_a} \quad (3-34)$$

Thus, we see that the deflection obtained in any cathode-ray oscillograph is directly proportional to the voltage applied to the deflecting plates and inversely proportional to the energy of the electron beam measured in volts, the proportionality factor depending only on the geometry of the tube.

3-5. The Magnetic Deflecting Type Oscillograph.—It sometimes happens that we are more interested in measuring current with an oscillograph than voltage or that it is not convenient to use voltage for deflecting the beam. In such a case, if we pass the current to be measured through small coils placed on opposite sides of the tube near the voltage deflection plates, it is possible to obtain deflections proportional to the current flowing in the

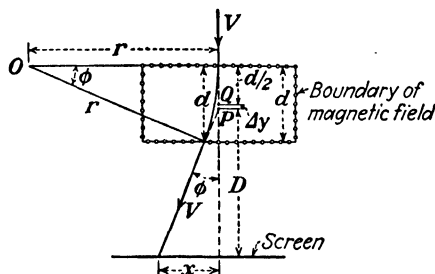


FIG. 3-8.—Electromagnetic deflection of cathode-ray beam.

coil. To cause deflections of the electron beam in this case, we make use of the magnetic field produced when the current passes through a coil. Since the force acting on an individual electron in the beam is proportional to the magnetic flux density and since, for an air-core coil, the magnetic flux density is directly proportional to the magnetic intensity or magnetizing force, the force on the electron will be directly proportional to the current flowing in the coil. The deflection observed on the screen will then be proportional to the current in the coil. This can be calculated as follows:

Assume that the magnetic field is of length d along the axis of the cathode-ray beam and, as shown in Fig. 3-8, is perpendicular to the plane of the paper. The deflection observed on the screen can be called x . If the magnetic field is uniform, the electrons will be turned through an angle ϕ while moving on the arc of a circle with center at O and radius r . The deflection observed on the screen is given by Eq. (3-35).

$$x = D \tan \phi \quad (3-35)$$

where D is the distance from P to the screen. Just as in the case of electrostatic deflections, P can be obtained by extending the trajectory of the electrons after they leave the deflecting field back to the intersection with the direction of the initial trajectory. In this case, however, P is not at the center of the magnetic field Q . The displacement of P from the center is so slight that for all practical purposes it can be considered to be at the center of the magnetic field.¹ Since the particle moves on a circular arc in the magnetic field, we have

$$\tan \phi = \frac{d}{\sqrt{r^2 - d^2}} = \frac{d}{r \sqrt{1 - (d/r)^2}} \quad (3-36)$$

The radius of curvature r is given by Eq. (3-18)

$$r = \frac{mV}{Bq_e}$$

and the velocity of the particles is given by Eq. (3-7) to be

$$V = \sqrt{\frac{2q_e E_a}{m}}$$

Therefore,

$$\tan \phi = d \sqrt{\frac{q_e}{2m}} \frac{B}{\sqrt{E_a}} \frac{1}{\sqrt{1 - (d^2 q_e B^2 / 2m E_a)}} \quad (3-37)$$

and

$$x = Dd \sqrt{\frac{q_e}{2m}} \frac{B}{\sqrt{E_a}} \frac{1}{\sqrt{1 - (d^2 q_e B^2 / 2m E_a)}}$$

For small values of B or large values of r , *i.e.*, for r greater than $10d$, we have, except for an error of less than $\frac{1}{4}$ per cent,

$$x = Dd \sqrt{\frac{q_e}{2m}} \frac{B}{\sqrt{E_a}} \quad (3-38)$$

¹ The distance Δy that P is displaced from the center of the magnetic field is given by a series of the form

$$\Delta y = \frac{d}{8} \left[\frac{d^2}{r^2} + \frac{1}{2} \frac{d^4}{r^4} + \frac{5}{16} \frac{d^6}{r^6} + \cdots \right]$$

To a first approximation, we have then

$$\Delta y = \frac{d^3}{8r^2} = \frac{d^3 B^2 q}{16m E_a}$$

If μ_0 is the permeability of free space,

$$B = \mu_0 H \quad (3-39)$$

and

$$H = \alpha NI \quad (3-40)$$

Here, α is simply a proportionality factor expressing the relationship between the magnetic intensity and the number of ampere turns produced in the coil. α will depend only on the shape of the deflecting coils. Combining Eqs. (3-38), (3-39), and (3-40), we see that the deflection produced on the screen is approximately given by

$$x = D d\mu_0 \alpha \sqrt{\frac{q_e}{2m}} \frac{NI}{\sqrt{E_a}} = K_2 \frac{NI}{\sqrt{E_a}} \quad (3-41)$$

We thus see that the deflection observed on the screen is directly proportional to the current in the deflecting coil and inversely proportional to the square root of the accelerating voltage.

It is interesting to compare this equation with Eq. (3-34) which gives the deflection in an electrostatic type oscillograph. The two expressions are similar in that the deflection of the beam is directly proportional to the voltage applied to the deflecting plates and the current flowing in the deflecting coil. But the two equations differ in their dependence on the accelerating voltage. Since in the electromagnetic oscillograph the beam deflection is inversely proportional to the square root of the accelerating voltage, changes in the accelerating voltage will affect the beam much less in this oscillograph than in the electrostatic type where the deflection is inversely proportional to the first power of the accelerating voltage.

3-6. Effect of Relativity Mass Correction.—With very high accelerating voltages the velocity of electrons in the cathode-ray beam approaches the velocity of light. In these circumstances the formulas above must be modified slightly, for the mass of the electron is really not a constant. It has been found experimentally that the actual mass of the electron increases with its velocity in such a way that it is impossible, even under the highest accelerating potentials, to make the velocity of electrons exceed or even equal the velocity of light. This is in agreement with the so-called "theory of relativity."¹ According to this theory the mass of the electron depends upon the velocity with which it is traveling, in the following way:

¹ See general reference 3-6.

$$m = \frac{m_0}{\sqrt{1 - (V^2/c^2)}} \quad (3-42)$$

In this equation, m_0 is the *rest mass* of the electron and is a universal constant and c is the velocity of light in vacuum. As can readily be seen, the effective mass of the electron differs only slightly from the rest mass when the electron is traveling at moderate velocities but approaches infinity as the electron velocity approaches the velocity of light. A simple calculation shows that if the accelerating voltage is 50,000 volts the effective mass of the electron is approximately 10 per cent greater than the rest mass and if the accelerating voltage is 500,000 volts the effective mass is approximately 100 per cent greater than the rest mass. This *effective mass* must be the one used in any accurate calculations of the motions of high-speed electrons. Thus, in the transverse field encountered by an electron of the beam in passing through the electrostatic field of the oscillograph deflecting plates, the force acting is given by

$$f = \frac{m_0}{\sqrt{1 - (V^2/c^2)}} \frac{dV_t}{dt} = qF \quad (3-43)$$

In this equation, V_t is the transverse velocity and V is the total velocity due principally to the accelerating potential and is not essentially changed by small transverse accelerations. Thus, it can be said that the effective mass appearing in Eq. (3-43) does not depend on the transverse velocity.

The mass does, however, depend upon the velocity of the particle when it is in the accelerating field at the cathode-ray-beam source, for here the electrons are being accelerated to large velocities in the direction of the field. The final velocities of the electrons in this field are dependent upon the increase in mass of the particles. The actual velocity of an electron in a large accelerating field can be calculated as follows: The force acting on any particle in an electric field is equal to the time rate of change of momentum of the particle by Newton's second law.

$$f = \frac{d}{dt}(mV) \quad (3-44)$$

Thus,

$$\begin{aligned} f &= m \frac{dV}{dt} + V \frac{dm}{dt} \\ &= m \frac{dV}{dt} + V \frac{dm}{dV} \frac{dV}{dt} \end{aligned} \quad (3-45)$$

If the mass of a particle is independent of its velocity, the second term on the right side of this equation drops out and we have the usual expression for the force in terms of the mass and acceleration. However, when the mass is dependent upon the velocity, all terms in the equation must be retained. It follows then that

$$q_e F = m \frac{dV}{dt} + V \frac{dm}{dV} \frac{dV}{dt} \quad (3-46)$$

Multiplying this equation by ds , we have

$$q_e F ds = m \frac{dV}{dt} ds + V \frac{dm}{dV} \frac{dV}{dt} ds$$

which can be rewritten

$$\begin{aligned} q_e F ds &= m \frac{ds}{dt} dV + V \frac{dm}{dV} \frac{ds}{dt} dV \\ &= mV dV + V^2 \frac{dm}{dV} dV \end{aligned} \quad (3-47)$$

If this equation is integrated along the path of acceleration, we have the energy gained by the particle in falling through a potential of E_a volts, or

$$q_e E_a = \int_{V=0}^{V=V_a} \left(mV dV + V^2 \frac{dm}{dV} dV \right)$$

Upon putting in the effective value of m , this equation becomes

$$q_e E_a = \int_{V=0}^{V=V_a} \left[\frac{m_0 V}{\sqrt{1 - \frac{V^2}{c^2}}} + m_0 V^2 \frac{\frac{V}{c^2}}{\left(1 - \frac{V^2}{c^2}\right)^{\frac{3}{2}}} \right] dV$$

This can be simplified by putting the integrand over a common denominator; then,

$$q_e E_a = \int_{V=0}^{V=V_a} \frac{m_0 V}{\left(1 - \frac{V^2}{c^2}\right)^{\frac{3}{2}}} dV \quad (3-48)$$

Integrating,

$$q_e E_a = \left[\frac{m_0 c^2}{\left(1 - \frac{V^2}{c^2}\right)^{\frac{1}{2}}} \right]_{V=0}^{V=V_a}$$

Upon putting in the limits, this becomes, since V_a is the velocity attained in falling through a potential E_a ,

$$q_e E_a = m_0 c^2 \left[\frac{1}{\left(1 - \frac{V_a^2}{c^2}\right)^{\frac{1}{2}}} - 1 \right] \quad (3-49)$$

By solving this equation for V_a , we can determine its dependence on E_a . We find that

$$V_a = c \left[1 - \frac{1}{\left(1 + \frac{q_e E_a}{m_0 c^2}\right)^2} \right]^{\frac{1}{2}} \quad (3-50)$$

or

$$V_a = \sqrt{\frac{q_e E_a}{m_0}} \left(\frac{\sqrt{2 + \frac{q_e E_a}{m_0 c^2}}}{1 + \frac{q_e E_a}{m_0 c^2}} \right) \quad (3-51)$$

We notice here, as well as in all previous equations, that if E_a is small or if V_a is small compared with the velocity of light the equation reduces to the one previously derived on the assumption that the mass was constant.

The ratios $\frac{m_0}{\left(1 - \frac{V^2}{c^2}\right)^{\frac{1}{2}}}$ and $\frac{m_0}{\left(1 - \frac{V^2}{c^2}\right)^{\frac{3}{2}}}$ appearing in Eq. (3-43) and

Eq. (3-48) are designated, respectively, as the *transverse mass* and the *longitudinal mass* of the electron; for the first arose from a consideration of the motion of the particle in a transverse field, and the second from the consideration of motion only in the direction of the field. Upon taking into account all effects of relativity, the deflection of an electron beam on a screen can be calculated as a function of the transverse and accelerating voltages. If we use the same procedure that was used in obtaining Eq. (3-34) but use for V_v the value of transverse velocity obtained from

$$V_v = at = \frac{q_e E_d}{bm_0} \sqrt{1 - \frac{V^2}{c^2}} t \quad (3-52)$$

and the value of V_a obtained from Eq. (3-51),

$$\begin{aligned} y &= D \frac{l}{b} \frac{E_d}{E_a} \frac{\left(1 + \frac{q_e E_a}{m_0 c^2}\right)}{\left(2 + \frac{q_e E_a}{m_0 c^2}\right)} \\ &= 2K_1 \frac{E_d}{E_a} \frac{\left(1 + \frac{q_e E_a}{m_0 c^2}\right)}{\left(2 + \frac{q_e E_a}{m_0 c^2}\right)} \end{aligned} \quad (3-53)$$

In this equation, y is the deflection on the screen, l is the length of the deflecting plate, and c is the velocity of light. All other symbols have the definitions used before. It is to be noticed that this equation reduces to Eq. (3-34) when the accelerating voltage is small.

To obtain some idea of the magnitude of this correction term for a typical high-voltage oscillograph, assume that E_a is 50,000 volts; then, $q_e E_a / m_0 c^2$ is approximately equal to 0.1. The correction term is

$$\frac{2 \left(1 + \frac{q_e E_a}{m_0 c^2}\right)}{\left(2 + \frac{q_e E_a}{m_0 c^2}\right)} = \frac{2.2}{2.1} = 1.05$$

We see, then, that the effect of the increased electron mass in a 50,000-volt oscillograph is to increase the deflections on the screen by about 5 per cent over the value that would be expected if the electron mass remained constant.

3-7. Application of the Oscillograph.—Ordinarily, observations made by means of the cathode-ray oscillograph of voltage alone or current alone are not very useful. For instance, if the

deflection plates of the oscillograph are connected to a source of alternating e.m.f., all that we observe on the screen is a line whose length is proportional to the maximum alternating voltage. Usually, we are interested in obtaining two-dimensional diagrams that give, for example, the dependence of voltage on time or the dependence of voltage on current. In any case, we want to observe the deflection of the beam when two forces are acting upon the beam at the same time. For this purpose, most oscillographs are made with two sets of deflecting plates set at

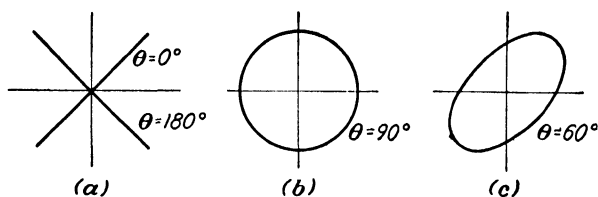


FIG. 3-9.

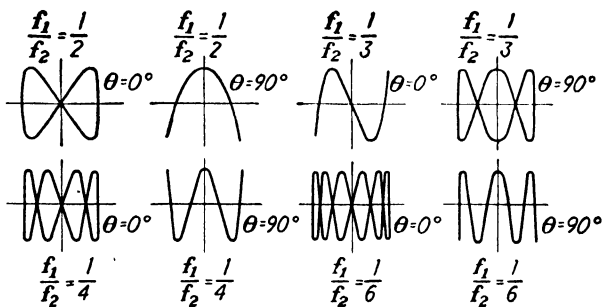


FIG. 3-10.—Lissajous figures.

right angles to each other. Thus, potentials between one pair of plates cause deflections in one direction, and potentials between the other plates cause deflections in a perpendicular direction. With this arrangement, it is possible to compare simultaneous values of two different voltages. As an example, suppose we wished to compare the frequencies of two sources. We could adjust the voltage applied to each set of plates until they were equal. Then, if the frequencies were the same the figure observed would be a 45-deg. line, a circle, or an ellipse, depending on the phase relationship between the two voltages, as indicated in Fig. 3-9a, b, and c, where θ is the phase difference between the two voltages. If the frequencies are not the same,

Lissajous figures similar to those shown in Fig. 3-10 are obtained.¹ In these figures, f_1 is the frequency of voltage on the horizontal plates and f_2 is the frequency of the voltage on the vertical plates. Initially (at time $t = 0$), the two voltages are both zero ($\theta = 0$ deg.) or 90 deg. out of phase ($\theta = 90$ deg.).

If we desire to compare simultaneous values of voltage and current, this can be done by impressing the voltage to be measured on one pair of plates and passing the current to be measured through a coil whose axis is perpendicular to the voltage plates and the electron beam. Then the force on the beam due to the current in the coil is at right angles to the force on the beam due to the potential on the voltage plates. It is possible to make the

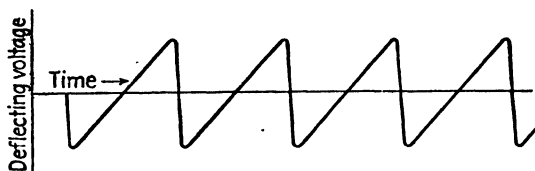


FIG. 3-11.—A.c. voltage for linear horizontal sweep of cathode-ray beam.

same measurements with two sets of voltage plates set at right angles if the current to be measured is allowed to pass through a resistance and the voltage drop in this resistance is impressed on the voltage plates. Voltage-current characteristics of this type are often desirable in studying arcs, glows, and the behavior of many kinds of high-vacuum tube.

In still other instances, we are interested in obtaining the dependence of voltage in a circuit on time, and for this purpose the set of plates giving a horizontal deflection can be energized with a voltage source that varies in some known way with the time. Usually, a horizontal deflection proportional to time is obtained by impressing an a.c. saw-toothed wave, such as is illustrated in Fig. 3-11, on the horizontal deflecting plates. A wave of this shape has the characteristics that, while the voltage is increasing, its amplitude increases linearly with time and, when the voltage decreases, it drops as rapidly as possible to its most negative value. Thus, when this wave is applied to the horizontal deflecting plates, the beam deflection increases linearly

¹ For other figures, see R.C.A. Pamphlet, *Technical Series TS-2*, "Cathode Ray Tubes and Allied Types," 1935.

with time, for the deflection is proportional to the applied voltage; and when the voltage reaches its maximum value and returns to its most negative value, the beam moves quickly from one side of the screen to the other and again starts to move uniformly across the screen as the voltage builds up. By properly adjusting the frequency of this deflecting voltage, a periodically recurrent voltage phenomenon applied to the vertical plates can be made to appear stationary on the fluorescent screen.¹

Although the linear sweep described above is the sweep encountered most frequently at the present time, it is not the only means of obtaining voltage-time or current-time oscillograms. In the study of most transient phenomena, there is no periodicity with which the sweep circuit can be synchronized. For example, lightning disturbances on high-voltage transmission lines set up transients that occur only over widely spaced intervals. In studying transients of this nature, it is customary to use a high-voltage cathode-ray oscillograph and record the transient directly on a photographic film. A condenser-charging or -discharging circuit is then used for the timing deflection and the circuit so adjusted by means of spark gaps or electronic tubes that the transient disturbance itself sets off the sweep circuit.

There is still another type of linear sweep sometimes used in high-voltage cathode-ray oscillographs that have the photographic film mounted inside the vacuum chamber. In these machines the film holder is made as an electrically driven rotating drum. One set of plates is all that is required with this arrangement, for the timing deflection is obtained by rapidly rotating the drum on which the film is mounted.

3-8. Focusing of Electron Beams.—Because of the fact that a cathode-ray beam consists of many electrons moving together, there is always a certain amount of spreading of the beam due to the mutual repulsion of the electrons. There are three methods in general use today to overcome this dispersion and make it possible to focus a cathode-ray beam to a small spot on the fluorescent screen. These can be called *gas focusing*, *magnetic focusing*, and *electrostatic focusing*.

¹ For the actual construction of cathode ray tubes and sweep circuits, see general reference 3-4 at the end of this chapter and K. Henney, "Electron Tubes in Industry," pp. 507-520, McGraw-Hill Book Company, Inc., New York, 1937.

The first method consists in introducing a small amount of inert gas into the cathode-ray tube, care being taken not to introduce so much gas that the electrons of the beam are appreciably scattered. With a slight amount of gas present a few of the electrons in the beam will make collisions with gas molecules before hitting the fluorescent screen. Each ionizing collision that a high-speed electron makes will introduce a positive molecular ion and an electron into the electron beam; but, because of the low energy and small mass of the new electrons, they will be quickly repelled toward the walls in the electrostatic field of the beam. In a very short time, enough positive ions will accumulate along and in the path of the beam actually to make the space charge positive instead of negative. Thus, instead of being repelled from the axis of the beam by the negative space charge

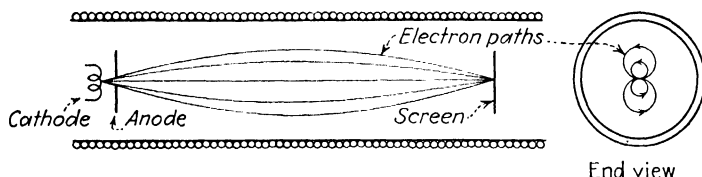


FIG. 3-12.—Focusing of a cathode-ray beam in a uniform magnetic field.

of the beam itself, the electrons will be attracted toward the center of the beam by the space charge of the positive ions they have produced. In some low-voltage cathode-ray tubes, this method can be used very successfully to obtain a fairly small spot.

With the magnetic-focusing method, use is made of the fact that the period of rotation of an electron in a uniform magnetic field is independent of the electron velocity and depends only on the magnetic field strength [see Eq. (3-23); the period, of course, is given by 2π divided by the angular velocity]. If the electron beam is sent down the axis of a long solenoid as shown in Fig. 3-12, those electrons whose path remains parallel to the axis of the solenoid will be undeflected but those which start to diverge will travel on helical paths and in one revolution be brought back to the axis. Regardless of the radial component of velocity that any electron may initially have, since they are all moving with essentially the same longitudinal component of velocity, all the electrons will be brought back to the axis at the same point. In most apparatus such as the cathode-ray oscillograph, it is not

practicable to use a long solenoid for focusing. It has been found, however, that it is possible to focus the beam in much the same way even when a relatively short coil is used. The important thing is to have the axis of the coil coincide with the axis of the electron beam.

At present, electrostatic focusing is used in most oscilloscopes in preference to the other methods. This method depends on the fact that the lines of force between two coaxial cylinders maintained at a difference of potential are not parallel to the electron beam. Figure 3-13 illustrates one arrangement. The

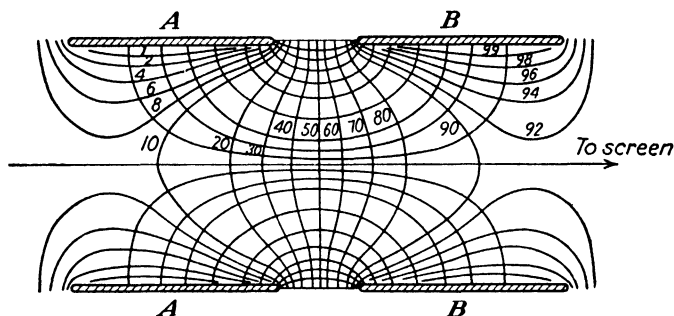


FIG. 3-13.—Electric field between two coaxial cylinders (potential difference—100 volts).

two electrodes *A* and *B* are to be imagined as rings down the axis of which the electron beam passes (see also Fig. 3-5b). The field map for this case has also been sketched in Fig. 3-13, on the assumption that the potential difference between the rings is 100 volts. We see, then, that as the electrons move toward the screen they are continuously accelerated; but, in addition, those electrons which tend to diverge from the axis of the cylinders find themselves in a radial electric field which gives them a component of velocity back toward the center. One might suppose that the diverging field near *B* would counteract the effect of the converging field near *A*; but actually the effect of the diverging field is less important than that of the converging field, for the electrons have gained energy in going from *A* to *B*. With this arrangement a very definite focusing action is possible. This arrangement of two coaxial cylinders maintained at a difference of potential is sometimes spoken of as an electron lens.¹

¹ ZWORYKIN, V. K., "On Electron Optics," *J. Franklin Inst.*, **215**, 535, 1933.

3-9. Anode Rays.—Beams of positively charged particles, anode rays, positive rays, or canal rays, are not encountered so frequently as cathode rays; but they are encountered in some problems and deserve consideration. Probably the most important applications of anode rays at the present time occur in studies of the nucleus of the atom, in the production of neutron beams, and in the production of artificially radioactive substances. It is not inconceivable that the engineer will soon be making as much use of the highly penetrating radiations from artificially radioactive substances as he has of the X ray. Thirty or forty years ago, the X ray was a tool of interest to the physicist alone for the examination of the structure of materials; but, with the passage of time, it has been found of great use in medicine and engineering as well as in physics. Similarly, we can expect that in the not too distant future the engineer will be finding uses for the highly penetrating radiations obtained from natural and artificially radioactive substances.

Anode rays are usually obtained by drawing off, by means of an electric field, positively charged particles of the gas in a glow- or arc-discharge tube. These positive ions are accelerated in an electric field and pass through slits or holes to define the beam just as electrons are accelerated in a cathode-ray tube. However, for producing transmutation of one element into another that is radioactive, the energy of the bombarding particles—the positive rays—must be very great. The particles must have kinetic energies of 1 million or more electron volts. There are several methods in general use for accelerating positive ions and giving them these high kinetic energies. The two principal methods are by means of the magnetic resonance accelerator, or cyclotron, and the Van der Graaf electrostatic generator.¹

Figure 3-14 shows a cross section of one type of apparatus for the production of high-energy positive rays.² This consists of a modified Van der Graaf high-voltage generator and a vacuum

¹ For a discussion of the experimental methods used in the production of high-energy positive ions the reader is referred to the very comprehensive article by M. S. Livingston and H. A. Bethe on "Experimental Nuclear Dynamics," *Rev. Mod. Phys.*, **9**, 246–260, 1937. This article, the first few pages of which are here referred to, contains also a very complete bibliography from which references to any type of accelerator can be obtained.

² HERB, R. G., C. M. TURNER, C. M. HUDSON, and R. E. WARREN, *Phys. Rev.*, **58**, 579, 1940.

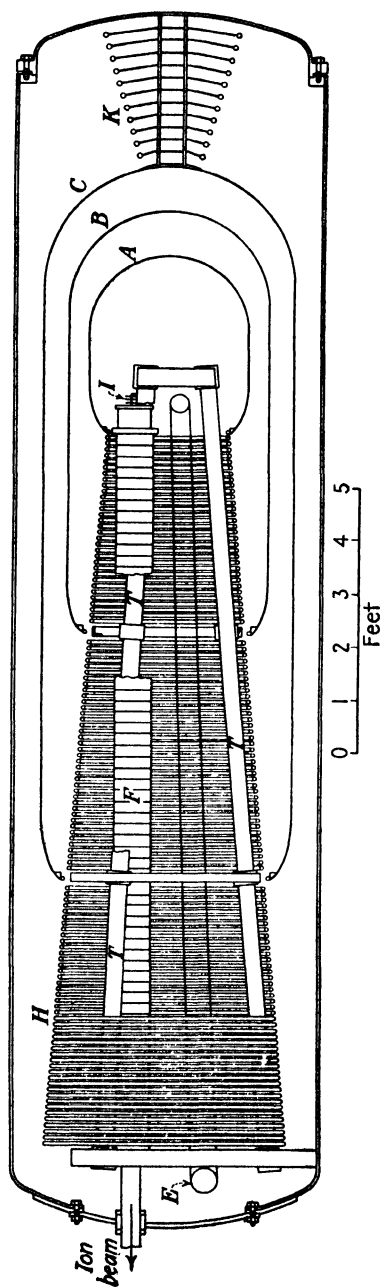


FIG. 3-14.—Van der Graaf high-voltage ion accelerator.

tube for the acceleration of the positive ions. The whole apparatus is mounted in a large steel tank and operated under about 100 lb./sq. in. of air pressure. The fact that the whole apparatus is placed in a high-pressure chamber makes it possible—as we shall see in a later section—to have much smaller spacings between the high- and low-potential elements than would otherwise be possible.

The innermost electrode *A* is raised to a high potential relative to ground by mechanically transferring charge from ground to *A* on an endless cotton fabric belt *E*. At one end of the belt a set of needle points is arranged so that, when a corona discharge is formed at the points, charges are sprayed onto the belt. At the other end of the belt, another set of needle points picks up this charge and transfers it to the high-voltage electrode. This electrode contains the discharge tube *I*, which serves as the source of positive ions. The accelerating tube *F* extends from the high-voltage electrode to ground and consists of 62 coaxial insulated metallic cylinders. The difference of potential between any two successive cylinders is adjusted to be approximately constant by means of a needle gap and a corona discharge outside the vacuum chamber. These corona gaps together with the hoops *H* serve to maintain a uniform potential drop along the tube and the insulating supports *T* for the high-voltage electrode. Thus, with a uniform potential drop along the vacuum tube the ions are uniformly accelerated and attain a velocity characteristic of the total potential fall when they reach the grounded end of the tube. Besides serving to accelerate the positive ions, the coaxial cylinders through which the ions pass keep the beam of positive ions focused throughout the whole length of the accelerating tube. The intermediate electrodes *B* and *C*, which are connected to the corona-gap voltage-distribution system, are used to maintain a nearly uniform voltage distribution in the space between the high-voltage electrode and the tank walls. With apparatus such as this, it is possible to obtain positive ions of hydrogen with kinetic energies of nearly 5 million volts.

The operation of the magnetic resonance accelerator, or cyclotron, is based on the fact that in a uniform magnetic field a charged particle moves on a circular path with an angular velocity which is determined solely by the strength of the magnetic field and the mass of the particle [see Eq. (3-23)]. Thus,

if a particle is projected into a uniform magnetic field, the time necessary for the particle to trace out a semicircular path is independent of the kinetic energy of the particle. The radius of its path, however, depends upon its energy, according to Eq. (3-18). In the case of the cyclotron, a source of electrons or positive ions is located at or close to the center of a uniform magnetic field. The ions, then, on entering a vacuum chamber placed in the magnetic field tend to travel on circular paths. Auxiliary electrodes, however, are placed in the vacuum chamber in such a way that every time the particle is turned through 180 deg. its energy is increased by a definite amount. This is accomplished by applying an alternating potential of a frequency equal to $qB/2\pi m$ between two pairs of D-shaped electrodes located with their straight sides adjacent and forming the two halves of a circle. Thus, every time an ion passes from between one pair of D's to the space between the other pair, it is accelerated by the amount of the difference in potential between the pairs of D's. During the time it is tracing out a semicircular path between one pair of plates the potential is reversing; and when it is ready to move back to the space between the other pair of plates, it is given another increment in energy. The particles, then, spiral out with ever-increasing energy until they leave the magnetic field or strike a target. Actually, only moderate voltages are needed with this apparatus in order to generate ion beams of several million electron volts energy, but the frequencies necessary are of the order of millions of cycles per second.

3-10. Television.—One very important application of the properties of cathode-ray beams, other than in the oscillograph, is in television apparatus.¹ Aside from the more or less mechanical means of televising, *i.e.*, by means of a rotating scanning disk, for example, there are two principal systems in use at present that make use of purely electronic methods of scanning. In one of these an electron image of the televised object is made to scan a collecting electrode, and in the other an electron beam scans a surface that has “memorized” an electronic image of the picture.

In the image dissector,² the picture to be televised is projected onto a photosensitive surface. Then, since such a surface emits

¹ For a good review of this subject, see *J. App. Phys.*, **10**, 426–493, 1939.

² Developed by Farnsworth Television, Inc.

more electrons from areas that are brightly illuminated than from dark areas, the density of electrons leaving the photoelectric cathode surface will vary from point to point in proportion to the illumination. These electrons are then accelerated in an electric field and focused by a magnetic field so that an electron image is formed in which the density of electrons in the beam cross section corresponds to the degree of shading—*i.e.*, lightness and darkness—in various parts of the picture. This electron image of the picture is then moved by means of crossed magnetic fields so that it sweeps across a scanning aperture 441 times as fast in the horizontal direction as in the vertical direction. Since the time required for a vertical sweep is $\frac{1}{30}$ sec., every element of the picture is thus scanned 30 times a second. Furthermore, the number of electrons collected at the scanning aperture is proportional to the brightness of the picture element. Thus, after suitable amplification the collector current can be broadcast on a radio wave and in a receiver can actuate the beam of a cathode-ray tube in one of the ways described below to reproduce the picture on a fluorescent screen.

TABLE 3-1*

Surface brilliance, candles/sq. ft.	Scene	Time	Date	Weather
700	Snow bank	10:00 A.M.	Jan. 24, 1935	Bright sunshine
500	Beach	2:00 P.M.	Aug. 18, 1934	Hazy
250	Bay	10:00 A.M.	June 30, 1935	Clear
130	Street	3:15 P.M.	June 30, 1935	Clear
40	Times Square	1:30 P.M.	Nov. 6, 1934	Light rain
55	Football game	1st quarter	Nov. 17, 1934	Clear
50	Football game	2d quarter	Nov. 17, 1934	Hazy
27	Football game	3d quarter	Nov. 17, 1934	Hazy
16	Football game	4th quarter	Nov. 17, 1934	Hazy
2	Football game	End	Nov. 17, 1934	Hazy
2	Open field	3:45 P.M.	July 1, 1935	Severe thunder storm

* Based on H. Iams, R. B. Jones, and W. H. Hickok, *Proc. I.R.E.*, **25**, 1034, 1937.

Because only about $1/250,000$ of the picture is over the scanning aperture at any instant, most of electrons emitted from the photosensitive surface never reach the collector. This means

that even though sufficient amplification is available the light intensities necessary for successful televising must be relatively high so that a clear picture can be obtained. Under conditions of low illumination, the random fluctuations in the electron emission from the photoelectric cathode are large in comparison with the total emission from any small area, and the televised picture

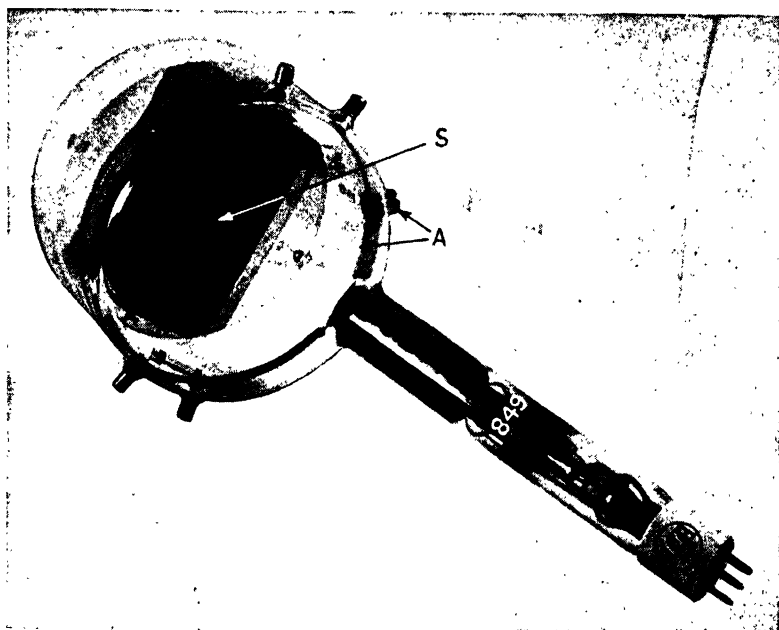


FIG. 3-15.—Iconoscope tube. S. Mosaic surface; A. Collecting electrode.
(Courtesy of R.C.A. Mfg. Co., Inc.)

presents a speckled appearance. In practice, it has been found that an average brightness of at least 150 to 170 candles/sq. ft. is necessary.

Although the iconoscope¹ is inherently more complicated than the image dissector, it has the advantage, at present, of being more sensitive under conditions of low illumination. Pictures can be picked up and transmitted by this tube even with an average brightness of 15 candles/sq. ft. Some notion of the meaning of these figures for surface brightness can be obtained from Table 3-1 shown on page 75.

¹ Developed by R.C.A. Mfg. Co., Inc.

The greater sensitivity of the iconoscope (Fig. 3-15) is due to the fact that in this tube each element of the surface S on which the picture to be televised is projected essentially "memorizes" its portion of the picture until a scanning electron beam produced in the tube passes by and transforms the element of the picture into a current pulse which can be sent out as a radio signal. Actually, the electrode onto which the pictures are projected consists of a mosaic of minute photosensitive particles of cesiated silver varying in size from 5×10^{-4} cm. in diameter down. These are spread over a thin mica sheet which in turn is backed by a solid metal plate. When a picture is projected onto the mosaic, the cesiated particles under light areas of the picture emit more electrons and assume a less negative potential than those in dark areas of the picture. Then, when the scanning electron beam passes and recharges all the particles to the same potential,¹ charges will be released from the small condenser formed by the cesiated particle and the backing plate at different rates in the lighted and unlighted areas. The change in current to the backing plate of the mosaic thus constitutes the picture signal.

Although this system has the advantage of greater sensitivity than the image dissector, it has the disadvantage that disturbances are easily introduced into the amplifier system by the redistribution of charges that is continually occurring in the mosaic. Furthermore, since under the condition of operation the maximum photoelectric emissivity cannot be utilized and since only about 25 per cent of those electrons which are emitted reach the collector electrode, the others returning to the mosaic, the actual efficiency of the tube is only a small fraction of its

¹ The particles are driven positive rather than negative by the scanning beam because each incident electron in the beam causes the emission of about 5 secondary electrons from the cesiated particles (see Sec. 6-10). The potential that they attain, however, is small because the secondary electrons emitted from the mosaic have a very low velocity and immediately return to the mosaic particle if its potential is more than 2 or 3 volts positive with respect to the collecting electrode (A , Fig. 3-15). This limits the positive potential that the particles can attain. Because of their low velocity, the emitted electrons form a space charge just outside the surface, and some of these electrons are continually returning to the mosaic. The particles of the mosaic thus gradually acquire a slight negative potential. In equilibrium the current to the collecting electrode is just equal to the scanning-beam current, and the potential of a mosaic particle fluctuates from a few volts positive to a few volts negative with respect to the collecting electrode.

theoretical efficiency. In practice, it has been found that surface brightnesses of 100 candles/sq. ft. are desirable, and there is thus no very great advantage in this system over that employing the image dissector.

Television receiving tubes differ only slightly from ordinary cathode-ray oscillograph tubes. In general, they consist of an electron gun and either an electrostatic or an electromagnetic means of producing deflections. By means of the deflecting plates—or coils—the cathode-ray beam is made to traverse a fluorescent screen in synchronism with the scanning beam in the transmitting tube. Then the picture itself is reproduced on the screen in one of three possible ways. In one method the beam moves across the screen with constant velocity, but the intensity of the beam is varied by means of the incoming signal to correspond to the shading in the picture. In another method the intensity of the beam is maintained constant, but its speed is varied by the incoming signal. In the third method, both the speed of scanning and the intensity of the beam are maintained constant, but the area of the spot on the fluorescent screen varies to produce the image.

Problems

1. Calculate the velocity of 25-volt electrons, of 25-volt singly charged hydrogen molecular ions, of 25-volt singly charged mercury ions. Mass of H_2 ion = 3.35×10^{-27} kg. Mass of Hg ion = 3.32×10^{-25} kg.

2. Given two infinitely extended parallel plates separated by a distance of 0.020 m. The potential of the upper plate relative to the lower is -400 volts. Electrons are shot out of the lower plate at the origin with an energy of 1,000 e.v. Where will these electrons strike the upper plate (or the lower plate if they do not hit the upper plate) if they are shot out at angles with the horizontal of (a) 30 deg., (b) 45 deg., (c) 60 deg., (d) 90 deg.? What will be the kinetic energy of the electrons when they strike the upper or lower plate? Use x - and y -components.

3. Solve Prob. 2 for the case in which the upper plate is at a potential of $+400$ volts relative to the lower.

4. Show that the path of the electron in Prob. 2a is a parabola.

5. The potential difference between the electrodes of an X-ray tube is 40,000 volts. If an electron starts from rest at the negative electrode, what will be its velocity when it strikes the positive electrode, if no collisions occur during its passage from the negative to the positive electrode? If the electron has an initial velocity of 6×10^7 m./sec., what will be its velocity upon reaching the positive electrode?

6. Electrons are shot into a uniform magnetic field of 0.01 weber/sq. m. with an energy of 2,500 e.v. What is the radius of curvature of the projection of the path on a plane perpendicular to the B vectors if the angle that the path makes with the B vectors is initially (a) 90 deg., (b) 60 deg., (c) 30 deg.?

7. In Prob. 6, if the electrons do not leave the magnetic field, what is the pitch of the helical path in each case?

8. In a magnetic field of 0.01 weber/sq. m. an electron traces out a circular path of radius 0.1 m. What is the initial energy of the electron? What is its energy after one revolution on the path?

9. How many r.p.m. would a singly charged hydrogen ion of 1,000 e.v. energy make in a magnetic field of 0.05 weber/sq. m.? Mass of H_2 ion = 3.35×10^{-27} kg.

10. A cathode-ray oscillograph has deflecting plates 1.0 cm. square and separated a distance of 5 mm. The electron beam consists of 800-volt electrons, and the distance from the deflecting plates to the fluorescent screen is 25 cm. What value of deflecting voltage is just sufficient to make the electron beam hit the lower edge of the plates? What is the maximum deflecting voltage that could be used if the screen is 15 cm. in diameter and the spot is to remain on the screen?

11. A certain cathode-ray oscillograph in which the accelerating voltage is 50,000 volts has a voltage sensitivity of 400 volts/cm. or a current sensitivity of 200 ma./cm., depending on the type of deflection used. Without changing the geometry of the tube itself (a) how could the voltage sensitivity be increased to 200 volts/cm.? (b) In what two ways could the current sensitivity be increased to 100 ma./cm.?

12. With a certain cathode-ray oscillograph having an accelerating voltage of 700 volts and deflecting coils 5 cm. in diameter, the observed deflection on the screen is 3 cm. What is the approximate flux density in the neighborhood of the electron beam if the distance from the deflecting coil to the screen is 0.15 m.?

13. Taking relativity corrections into account, what is the actual velocity of the electrons described in Prob. 5? What is the mass of the electrons when they strike the positive electrode?

14. What is the mass of an electron that has been accelerated through 1 million volts? What is its velocity?

15. Construct the Lissajous figure that results when a 90-cycle voltage is applied to the vertical deflecting plates of a cathode-ray oscillograph and a 60-cycle voltage is applied to the horizontal deflecting plates. Assume that the peak value of each voltage wave is the same and that both voltage waves are initially zero.

General References

- 3-1. Dow, W. G.: "Fundamentals of Engineering Electronics," pp. 56-96, John Wiley & Sons, Inc., New York, 1937.
- 3-2. CROWTHER, J. A.: "Ions, Electrons, and Ionizing Radiations," 7th ed., Chap. VII, Edward Arnold & Co., London, 1938.

- 3-3. THOMSON, J. J., and G. P. THOMSON: "Conduction of Electricity through Gases," Vol. I, Chaps. V and VI, pp. 215-290, Cambridge University Press, London.
- 3-4. MACGREGOR-MORRIS, J. T., and J. A. HENLEY: "Cathode Ray Oscillography," Chapman & Hall, Ltd., London, 1936.
- 3-5. MALOFF, I. G., and D. W. EPSTEIN: "Electron Optics in Television," McGraw-Hill Book Company, Inc., New York, 1938.
- 3-6. PAGE, LEIGH: "Introduction to Theoretical Physics," pp. 518-530, D. Van Nostrand Company, Inc., New York, 1935.

CHAPTER IV

THE MOLECULAR THEORY OF GASES¹

4-1. Introduction.—A wide range of conditions of the gas is encountered in dealing with problems of electrical conduction in gases. The pressure may be so low, as in a high-vacuum radio tube, that the gas plays little part in the electrical conduction through the tube. In other cases a pressure of a few millimeters of mercury is encountered, and the conduction takes place by virtue of the presence of swarms of ions intermingling with the much larger numbers of neutral gas molecules. Electrical discharges occurring in practice also exist under high-pressure conditions. An example is the arc that forms as the contacts separate in a large oil circuit breaker. The heat developed by the arc current decomposes some of the oil in which the arc is immersed, giving gaseous end products which form a bubble or gas pocket in which the arc burns. If the circuit breaker is called upon to interrupt short-circuit currents of several thousand amperes, the pressure in the gas surrounding the arc may build up to over 10 atm. before the arc is extinguished.

In the quantitative treatment of these widely differing problems, the common feature consists in the motions of the electrically charged ions through the gas. The motion of a single ion through evacuated space can be predicted by the methods discussed in Chap. III, if the mass and charge of the ion and the electric intensity in the region are known. However, if the ion moves through a gas in which the pressure is appreciable, the ion is likely to have collisions with the gas molecules. These collisions will profoundly alter the motions of the ions in the electric field. Furthermore, the motions of the gas molecules, themselves, may have an effect on the resulting motions of the ions. In attacking problems of the motions of ions through gases, it will be found helpful to start with a knowledge of the mode of motion of the neutral molecules and then proceed to extend these concepts to the calculation of the ionic motions. The

¹ General references for Chap. IV will be found on p. 124.

present chapter will be devoted to a study of some of the more important topics relating to the motions of the molecules of a gas.

The early chemists Dalton and Berzelius had practically established the molecular theory of the constitution of matter by about 1810. The molecular theory as applied to gases, especially the calculation of pressure and internal energy on the basis of molecular motions, was worked out in 1857 by R. Clausius. The theory was expanded by J. Clerk Maxwell, Clausius, and Lord Kelvin in the period 1860 to 1880. Subsequently, valuable contributions were made by Boltzmann, Van der Waals, Tait, and O. E. Meyer. The careful studies of the motions of ions carried out by Langevin, J. J. Thomson, Townsend, and K. T. Compton have also served to extend the theory of molecular motions. The earliest direct experimental observations that proved the existence of the heat motions of the molecules were reported by Perrin¹ in 1908. Many more recent experiments showing agreement between the theory and experiment have been carried out in studies of heat conduction, the viscosity of gases, and the conduction of electricity through gases.

4-2. Experimental Basis of Gas Theory: The General Gas Law for Ideal Gases.—The experimental work of Boyle, Charles, and Gay-Lussac led to experimentally determined relations between the pressure, volume, and temperature of a given mass of gas that can be expressed by the general gas equation

$$pv = R'T \quad (4-1)$$

In this equation, p is the pressure, v is the volume, R' is a constant depending on the mass and kind of gas, and T is the absolute temperature, *i.e.*, the temperature measured on the Kelvin thermodynamic scale. This equation is rigorously valid only for a so-called "ideal" or "perfect" gas but is sufficiently accurate for real gases that are not too near their condensation temperature. The further a real gas is from its condensation temperature, the more nearly Eq. (4-1) will portray its actual behavior.

¹ Perrin studied the Brownian movements of very small particles suspended in a liquid. J. Perrin, "Atoms," 2d English ed., Constable & Company, Ltd., London, 1923. See also R. A. Millikan, "Electrons (+ and -), Protons, Photons, Neutrons, and Cosmic Rays," Chap. VII, University of Chicago Press, Chicago, 1935, for an account of studies of the Brownian movements of small particles in gases.

If we are dealing with a mass of gas equal to the molecular weight¹ Eq. (4-1) may be written,

$$pv = RT \quad (4-2)$$

in which R is independent of the kind of gas and is called the *molar gas constant*. The value of this constant is

$$R = 8.3136 \pm 0.001 \text{ joule per degree Kelvin} \quad (4-3)$$

If a mass of gas other than that equal to the molecular weight is considered, the pv product will, of course, be changed proportionately:

$$pv = \frac{M}{M_0} RT \quad (4-2a)$$

in which M is the mass being considered and M_0 is the mass of a molecular weight of the gas.

If Eq. (4-2) is considered from the viewpoint of dimensional analysis, we find that the left side represents work or energy. This follows because the pressure has the dimensions of force divided by a length squared whereas the volume has the dimensions of length cubed, and the product therefore has the dimensions of force multiplied by length. Consequently the right side, or RT , must also have the dimensions of energy. Moreover, since R is a constant, we see that the energy represented by the right side of the equation is proportional to the absolute temperature T . If we consider the case of a constant volume of gas v_0 as the temperature is varied, we may write

$$pv_0 = RT$$

As heat is supplied to the gas, its temperature rises, the pressure increases, and the gas has an increased capacity for doing work if allowed to expand. At constant volume the increase in pressure is in the same proportion as the increase in temperature and as the increase in energy. The solution of the problem of finding

¹ By the molecular weight of a gas is meant the sum of the atomic weights of the elements that form the gas molecule. An atomic weight may be defined as the relative mass of the element referred to a common standard. Though hydrogen was originally used for the standard of reference, the modern standard is oxygen, and its atomic weight is taken as 16 g. The molecular weights are usually expressed in grams, but in the m.k.s. system they will be expressed in kilograms as given in Appendix III.

a mechanism by virtue of which both the pressure and the energy of the gas would vary proportionately followed naturally from the assumption that the gas consisted of molecules in continual motion. Thus the pressure exerted by the gas is nothing more than the smoothed-over effect of the impacts of large numbers of molecules on the walls of the vessel, and the internal energy is the kinetic energy of the moving molecules. Furthermore, the temperature is directly proportional to the energy of the molecules. These ideas will be developed in a quantitative way in the following sections.

4-3. The Kinetic Picture of a Perfect Gas.—According to the molecular theory, the properties of gases, for example, the pressure, may be accounted for in a quantitative manner on the basis of a simple set of postulates. For an ideal gas these postulates are as follows:

1. The gas is considered to consist of small particles or molecules which obey the laws of the Newtonian mechanics.

2. The particles are considered to be perfectly elastic spheres and to be all alike for a given kind of gas.

3. The volume occupied by the molecules is postulated to be small compared with the volume of the vessel holding the quantity of gas in question.

4. The particles are conceived to be in ceaseless, chaotic motion such that the particles may have collisions with other particles and with the walls of the vessel.

5. It is furthermore assumed that the attractive or repulsive forces between the particles or between the particles and the walls are small except when a particle is in the immediate vicinity of another particle or the wall. Also, the forces of gravity are assumed to be negligible. As a result, the motions of the particles between successive impacts consist of rectilinear, translatory motions.

6. In any given volume the number of molecules is conceived to be very large.

7. Finally, when the distribution of molecular speeds is being considered, it will be necessary to assume that the molecules have been colliding for a long enough time so that an equilibrium chaotic condition has been reached.

It is seen that the expansive action of a gas follows naturally from the assumptions given above. In a qualitative way, it

can be seen that the pressure on the walls of the container is explained as caused by the net effect of the impacts of the molecules on the containing surfaces.

4-4. Elementary Calculation of Pressure and Molecular Speed.—A very simple derivation may be given for the pressure of a gas in terms of the motions of the molecules if an analysis is made on the basis of simplified assumptions.

Let there be n molecules of a given kind of gas contained in the cubical box having edges of length s as shown in Fig. 4-1. Although the individual molecules are actually moving with different velocities and in a random, chaotic fashion, it will simplify the analysis greatly to assume that the molecules are all moving with a certain average speed C , and furthermore to assume that on the average one-third of the molecules may be considered as moving parallel to each of the x -, y -, and z -axes. The molecules are assumed to traverse the whole path from one side of the box to the other without colliding with other molecules.

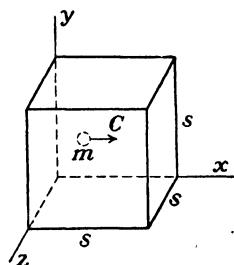


FIG. 4-1.—Cubical box containing n molecules, each of mass m moving with the speed C .

Consider a molecule of mass m which is moving in the positive x -direction and is approaching the right-hand face of the box with the speed C . Since the collision of the molecule with the wall is assumed to be perfectly elastic, there will be no energy lost, and the molecule will rebound with the same speed C . To bring about this change in the direction of the velocity of the molecule, the wall must exert a force f on the molecule during the time of contact¹ that is given by Newton's second law of motion, namely:

$$f = m \frac{dV}{dt} \quad (4-4)$$

Here V is the velocity of the molecule in the positive x -direction. According to Newton's third law the molecule must exert an equal but opposite force on the wall. It is proposed to calculate an average or smoothed-over force that will be equivalent to the effect of the impacts of all the molecules that strike the right-hand

¹ The time of contact for an actual molecule may be defined to be the total time during which forces are acting between the wall and the molecule.

wall. We first calculate the average force exerted by a single molecule. A schematic diagram of the manner in which the velocity of a molecule changes with time during the impact is shown in Fig. 4-2a. The corresponding variation of the force exerted on the wall is also shown. Now, let t_2 be the time required for the molecule to move back to the left-hand wall, rebound, and return to strike the right-hand wall, *i.e.*, the time for a round trip. Then there will be one impact on the right-

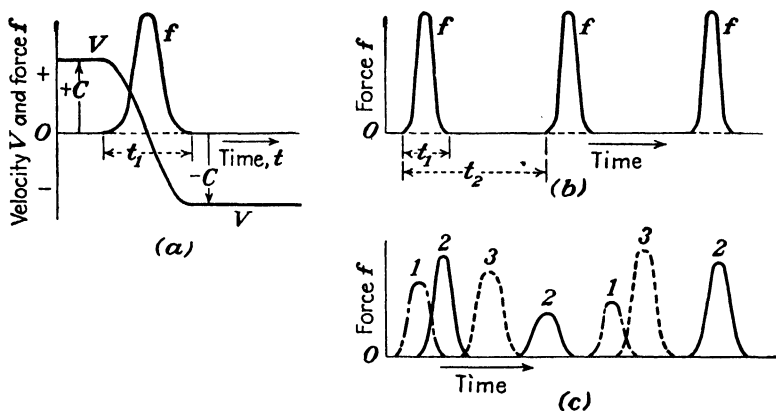


FIG. 4-2.—(a) Velocity and force curves for an impact of a molecule on a wall; (b) successive impacts of a molecule on a wall; (c) successive impacts of several molecules on a wall.

hand wall every t_2 sec., and there will be a succession of transient forces acting on the wall as shown in Fig. 4-2b. The average force f_a on the wall resulting from one impact will be

$$f_a = \frac{- \int_0^{t_1} m \frac{dV}{dt} dt}{t_2} = \frac{-m \int_C^{-C} dV}{t_2}$$

where t_1 is the time of contact. Thus,

$$f_a = \frac{2mC}{t_2} \quad (4-5)$$

Equation (4-5) also gives the average force on the wall caused by one molecule averaged over any length of time, provided that the time is long compared with t_2 . Therefore, the average force, f'_a , exerted by all the molecules moving in the x -direction (by assumption $\frac{1}{2}n$ in number) is given by the equation

$$f'_a = \frac{2mnC}{3t_2} \quad (4-6)$$

The time required to traverse twice the width of the box is given by

$$t_2 = \frac{2s}{C}$$

Substituting this value in Eq. (4-6), we obtain

$$f'_a = \frac{mnC^2}{3s}$$

The pressure is the force per unit area; therefore,

$$p = \frac{f'_a}{s^2} = \frac{mnC^2}{3s^3} \quad (4-7)$$

Since s^3 is the volume of the box, v , Eq. (4-7) may be rearranged to give

$$pv = \frac{1}{3}mnC^2 \quad (4-8)$$

This basic equation shows how the product of pressure by volume depends on the mass, number, and speed of the molecules. It is apparent that, if the speed of the molecules is constant at constant temperature, then the product of pressure and volume will also be constant, in agreement with Boyle's experimental law.

Though Eq. (4-8) was derived on the basis of simplified assumptions, it may be demonstrated¹ that a more rigorous treatment leads to the same result. In the actual gas the speeds of the molecules vary from time to time and from molecule to molecule, so that the impacts of the molecules on a wall would vary in frequency and intensity with a resulting schematic force diagram as shown in Fig. 4-2c. When we consider only a few molecules, as in the figure, the force on the wall fluctuates widely; but the molecules in any sample of gas are so very numerous that the total force is constant for all practical purposes. The interpretation of C and the question of the varying velocities will be discussed in detail in Secs. 4-8 to 4-11.

The average² speed C in Eq. (4-8) may be determined from experimental data on gas densities and pressures, for the density

¹ General reference 4-1, pp. 8-11.

² This average will later be shown to be the *root-mean-square* (r.m.s.) average speed (see Sec. 4-11).

of the gas will be the mass of all the molecules divided by the volume, *i.e.*, mn/v ; and, from Eq. (4-8), we obtain the expression

$$p = \frac{1}{3}dC^2 \quad (4-9)$$

Solving this expression for the molecular speed, we obtain

$$C = \sqrt{\frac{3p}{d}} \quad (4-10)$$

As an example, consider oxygen gas (O_2), for which $d = 1.43$ kg./cu. m. at a pressure of one standard atmosphere, *i.e.*, $p = 1.013 \times 10^5$ newtons/sq. m., and at a temperature of $0^\circ C$. Substituting these values in Eq. (4-10), we find that the speed C is 459.4 m./sec., or 1,507 ft./sec. This velocity is of the order of the muzzle velocity of a rifle bullet and is somewhat above the velocity of sound waves through air.

4-5. Avogadro's Hypothesis.—The development of the atomic theory by Dalton on the basis of the combining weights of the elements and the work of Gay-Lussac on the chemical combination of gases led up to the enunciation by Avogadro in 1811 of the hypothesis that bears his name. This may be stated as follows: *Under the same conditions of temperature and pressure equal volumes of all gases contain the same number of molecules.*¹ Avogadro's hypothesis was deduced from the fact that masses of different gases in proportion to their molecular weights occupy the same volume. Avogadro was the first to recognize that it was necessary for the idea of the molecule as a group of atoms to include the possibility of a group of atoms of the same element in order to clarify the experimental results on the chemical combination of gases.

The truth of this hypothesis is implied in the statement of the general gas law in the form of Eq. (4-2). In this equation, R is independent of the kind of gas if a molecular weight is considered. Thus, Eq. (4-2) states that, at any fixed reference pressure and temperature, a molecular weight of gas and therefore, also, a given number of molecules, will occupy a fixed volume. This volume turned out to be 22.41 liters if the reference temperature and pressure were $273^\circ K$. and one standard atmosphere, respectively, *i.e.*, under *standard conditions* of temperature and pressure.

¹ The word *molecule* must be construed to include the special case in which the gas particles are atoms, as in a monatomic gas.

The truth of Avogadro's hypothesis has been tested in many ways,¹ and it is now considered as an established law. The number of molecules N_0 in the 22.41 liters of gas under standard conditions has been called the *Avogadro number*. The value of this constant has been determined by several methods,² and its value is

$$N_0 = 6.023 \times 10^{23} \quad (4-11)$$

After the number of molecules in a molecular weight is known, it is a simple matter to calculate the number of molecules in a unit volume of gas under any conditions. The term *molecular concentration* will be used to denote the number of molecules per unit volume. Letting N symbolize the molecular concentration and considering a molecular weight of gas, we have, by definition,

$$N = \frac{N_0}{v} = \frac{N_0 p}{RT} \quad (4-12)$$

where v is obtained from Eq. (4-2). This general relation shows that the molecular concentration varies directly with the pressure and inversely as the temperature of the gas. For example, let us consider a fixed mass of gas contained in a cylinder having a tight-fitting piston. If the pressure on the gas is maintained constant while the temperature of the gas is increased, the piston will move outward, and the molecular concentration will decrease in accordance with Eq. (4-12). On the other hand, if the temperature is maintained constant and the pressure is increased, the piston will move inward, and the molecular concentration will increase in proportion to the increase in pressure.

The value of the molecular concentration under standard conditions is 2.69×10^{25} molecules/cu. m. It is almost impossible to conceive such a large number. If a very small volume is imagined, for example a cube 0.1 mm. on an edge, the number of molecules contained in it is still enormous, being 2.69×10^{13} for the volume mentioned. Suppose we calculate the molecular concentration by Eq. (4-12) for a relatively high vacuum, say a pressure of 10^{-7} mm. Hg, which is equivalent to

$$10^{-7} \times 1.013 \times \frac{10^6}{760} = 1.333 \times 10^{-8} \text{ newton/sq. m.}$$

¹ General reference 4-3, p. 24.

² For example, see R. A. Millikan, *op. cit.*, p. 120.

Let the temperature be $273^{\circ}\text{K}.$; then we have

$$N = \frac{6.023 \times 10^{23} \times 1.333 \times 10^{-5}}{8.314 \times 273} = 3.54 \times 10^{15} \text{ molecules/cu. m.}$$

Thus at this high-vacuum condition there are still 3.54 million molecules in a cubic millimeter of gas.

4-6. Temperature and Molecular Energy and Speed.—In Sec. 4-2 it was pointed out that the absolute temperature of a gas is directly proportional to the energy of the gas. Thus the absolute temperature may be looked on as being defined in terms of the energy of the gas. To deduce the relation between the temperature and energy of a gas on the basis of the molecular theory, we equate the right sides of Eqs. (4-2) and (4-8) and after replacing n by the number of molecules in a molecular weight N_0 we find

$$pv = \frac{1}{3}mN_0C^2 = RT \quad (4-13)$$

The kinetic energy W of the N_0 molecules moving with the speed C is $N_0 \times \frac{1}{2}mC^2$, and therefore Eq. (4-13) leads to the expression

$$W = \frac{3}{2}RT \quad (4-14)$$

Letting the kinetic energy of a *single* molecule be represented by W_k , we immediately obtain its value

$$W_k = \frac{3}{2} \frac{R}{N_0} T$$

The constant ratio R/N_0 appears so often in computations dealing with gases that it has been given a name, *i.e.*, Boltzmann's constant, and it is invariably represented by the symbol k .

$$k = \frac{R}{N_0} = \frac{8.314}{6.023 \times 10^{23}} = 1.380 \times 10^{-23} \text{ joule per degree Kelvin} \quad (4-15)$$

Upon using this constant, the kinetic energy of a single molecule may be expressed as

$$W_k = \frac{3}{2}kT \quad (4-16)$$

Thus, the kinetic energy of translation of a single molecule is directly proportional to the temperature of the gas.

To find the relation between the molecular speed and the temperature, we solve Eq. (4-13) for the speed; and, after substituting k for R/N_0 , we have

$$C = \sqrt{\frac{3kT}{m}} \quad (4-17)$$

We note that the average molecular speed varies as the square root of the absolute temperature and is independent of such variables as pressure and volume except inasmuch as these have an effect on the temperature of the gas.

It is worth noting that in the analysis above the energy that might be stored in the gas in forms other than kinetic energy of pure translatory motion has not been considered. For example, there is a possibility of energy of rotation of the molecule or of energy associated with oscillations of the atoms about their positions of equilibrium in the molecule. Rotational motions and oscillatory motions within the molecule play no part in producing pressure on the gas boundary and therefore need not be considered in our calculations. However, if we were concerned with the amount of heat energy required to raise the temperature of the gas, *i.e.*, if we were to attempt to calculate the specific heat of the gas, we should need to examine more closely the possible rotational and oscillatory motions of the molecules.

4-7. Mixed Gases.—Dalton's law of partial pressure, which resulted from his experimental investigations of the pressure relations for a mixture of different gases, is embodied in the following statement: *The pressure exerted by a mixture of gases is equal to the sum of the pressures that would be exerted by each of the individual gases if it alone were present in the same volume at the same temperature.* This relation may be expressed mathematically as follows:

$$p = p_1 + p_2 + p_3 + \cdots \quad (4-18)$$

Here the symbol p_1 refers to the pressure that the gas molecules of type 1 would exert if they alone were present; p_2 , p_3 , $\cdots$ similarly refer to the partial pressures exerted by the remaining types of molecules in the mixture; and p is the pressure in the mixture itself.

If the partial pressures in Eq. (4-18) are expressed in terms of the molecular speeds by the use of Eq. (4-8), the following

expression results:

$$p = \frac{m_1 n_1 C_1^2}{3v} + \frac{m_2 n_2 C_2^2}{3v} + \frac{m_3 n_3 C_3^2}{3v} + \dots \quad (4-19)$$

Furthermore, since n_1/v is the molecular concentration of the molecules of type 1 and $(\frac{1}{2})m_1 C_1^2$ is the kinetic energy $(W_k)_1$ of these molecules, Eq. (4-19) will yield the following expression for the total pressure in terms of the molecular concentrations and the kinetic energies of the molecules:

$$p = \frac{2}{3}N_1(W_k)_1 + \frac{2}{3}N_2(W_k)_2 + \frac{2}{3}N_3(W_k)_3 + \dots \quad (4-20)$$

However, Eq. (4-16) states that W_k is equal to $\frac{3}{2}kT$, showing that the kinetic energy of a molecule depends only on the temperature and not on the kind of molecule; thus, the kinetic energies of the different molecules of the mixture will be the same. Therefore, we drop the subscripts from the W_k 's in Eq. (4-20) to obtain

$$p = \frac{2}{3}W_k(N_1 + N_2 + N_3 + \dots) \quad (4-21)$$

This equation shows that the pressure of the mixture of gases varies directly with the sum of the molecular concentrations of the several gases.

The principle used in the derivation of Eq. (4-21), namely, that *the kinetic energies of the molecules of a mixture of gases in equilibrium are the same*, is worthy of additional emphasis. As an example of the application of this principle, consider the equilibrium between a mixture of a few free electrons with a gas. If we treat the free electrons as a perfect gas, we can readily find their velocity of thermal agitation. On the basis of the principle of equality of kinetic energies, we write the equation

$$\frac{1}{2}m_0 C_e^2 = \frac{1}{2}m C^2$$

Here, m_0 is the mass of the electron, C_e is its speed, and m and C are the mass and speed of the gas molecules. The value of C_e obtained on this basis is

$$C_e = \sqrt{\frac{m}{m_0}} C$$

Suppose we apply this expression to the calculation of the speed of the electrons in oxygen gas under standard conditions. The

mass m of the oxygen molecule equals the molecular weight divided by Avogadro's number, or

$$\frac{0.032}{6.023 \times 10^{23}} = 5.31 \times 10^{-26} \text{ kg.}$$

The speed of the oxygen molecule was calculated in Sec. 4-4 to be 459.4 m./sec. under the given conditions. The mass of the electron may be taken as 9.1×10^{-31} kg. Substituting in the formula, we find that C_e equals 111,000 m./sec., or about 242 times the speed of the oxygen molecules. The fact that the electron velocities are so high will be of use when we discuss the electron motions through a gas in Sec. 4-12.

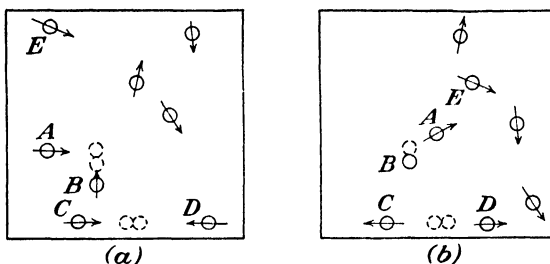


FIG. 4-3.—Molecular velocities.

4-8. Distribution of Molecular Velocities: Introduction.—In Sec. 4-4 the general gas law was derived on the assumption of uniform speeds for all the molecules in a given enclosed volume of gas. However, a consideration of the behavior of a large number of molecules in such an enclosed volume shows the impossibility of uniform speeds for all the molecules. Although the molecules are conceived to be small, the large number of molecules in a given volume (see Sec. 4-5) increases the likelihood of collisions between the molecules. Since these collisions are assumed to be perfectly elastic, the final velocities for a given impact could be calculated if the velocities before the impact and the masses of the molecules are known. The conditions at two successive moments in a gas might be pictured as in Fig. 4-3a and b if molecules whose velocity vectors lie in a plane are considered. The circles in Fig. 4-3 represent the molecules, and the vectors represent their velocities at the particular moment. The dotted circles in Fig. 4-3a indicate the positions the molecules will have when the next impact takes place; in Fig. 4-3b the dotted circles

show these same positions, but the collisions have already taken place. Two rather special cases of collisions are illustrated, the head-on collision of molecules *C* and *D*, and the collision of molecules *A* and *B*. *A* and *B* collide in such a way that *B* has zero velocity after the collision and *A* has a velocity equal to the vector sum of the velocities *A* and *B* had before the collision. In the actual case the velocities in a three-dimensional space must be considered, and the picture becomes more complicated. However, it is clear that at any instant a wide range of velocities from zero to ten or more times the average velocity will be found in the gas.

On account of the special nature of the collisions required to bring a molecule to rest, it can be inferred that relatively few molecules will have velocities at or near zero." A little reflection will show that not many molecules will have very high velocities, for a succession of just the right type of collision would be needed if the molecule's velocity is to increase at each collision. These qualitative arguments lead one to expect that most of the molecules will, at any instant, have speeds not far removed from the average speed and that fewer molecules will have the low speeds or very high speeds.

Maxwell was the first to derive an expression that described the probable ways in which the molecules in a quantity of gas are moving at any instant. He was able to deduce this relation without examining into the details of the collision processes. Boltzmann¹ treated the problem in a more rigorous fashion by studying the equilibrium conditions in a gas that result from the elastic impacts of the molecules. Before proceeding to give a brief form of Maxwell's derivation of the important velocity distribution laws, we digress to give an introductory example to illustrate the meaning and forms of distribution laws.

Let us imagine that we have been assigned to the task of reporting on the speeds of the automobiles that travel past a particular point on a highway. Suppose we gather data on the speed of each car passing the observation point for an entire day and find that we have data on 1,000 cars. Now the question arises: How can this data best be presented? A simple and vivid representation of the data would be a graphical one such as shown by curve *A* in Fig. 4-4a. This curve gives the *number*

¹ General reference 4-3, Chap. IV.

of cars in the different 10-m.p.h. speed ranges such as 0 to 10, 10 to 20, etc. as ordinates plotted against the speed as abscissas. Though curve *A* gives a clear representation of the "distribution of speeds" of the cars in 10-m.p.h. ranges, it is not fine-grained enough to yield accurate information about the number of cars that were traveling at speeds, say between 55 and 60 m.p.h. In order to answer questions similar to this relating to details of the distribution curve, it is desirable to have a continuous mathematical function that will represent the data on the speeds of the

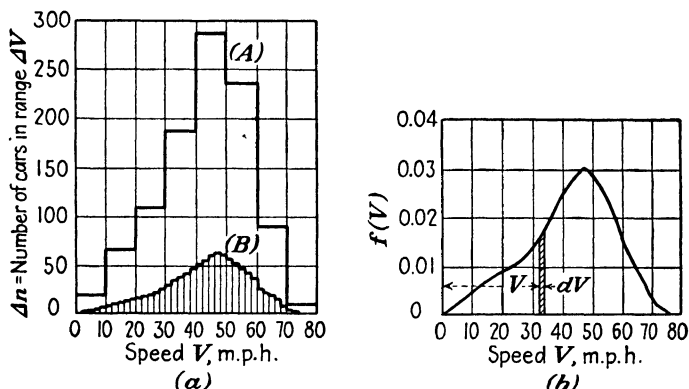


FIG. 4-4.—(a) Step type distribution curves. $\Delta V = 10$ m.p.h. for curve *A*, and $\Delta V = 2$ m.p.h. for curve *B*. (b) Continuous distribution curve.

cars. The first step in obtaining such a function would be to plot the distribution of speeds of the cars using smaller ranges of speed. Curve *B* in Fig. 4-4a shows such a curve using 2-m.p.h. ranges of speed. This shows that subdividing the original speed range has the effect of giving a smoother "step type" curve. We also note that the number of cars found in the different ranges is smaller on account of the smaller range of speeds. If we continue to subdivide the speed ranges, we should soon find that the resulting distribution curve would no longer continue to become smoother but instead would begin to have relatively higher percentage changes in ordinate from one range to the next. This would occur when the numbers of cars in the ranges become too small. For example, if the range of speeds is taken equal to 0.1 m.p.h., part of the data for the curve would appear as in Table 4-1. Thus, we must cease to subdivide the speed ranges before the number of cars in each range becomes too small.

TABLE 4-1

Speed Range, M.P.H.	Number of Cars in the Range
40.0-40.1	2
40.1-40.2	0
40.2-40.3	1
40.3-40.4	3
40.4-40.5	0
40.5-40.6	2

After obtaining the smoothest possible step type curve, we draw a smooth curve through the mid-points of the horizontal segments of the step type curve. Though we might regard the resulting smooth curve as representing the function for which we have been searching, we can raise one objection to its use, that the height of the curve depends on the magnitude of the range of speeds adopted for the best step type curve. This objection can be removed by the use of a more general function $f(V)$, defined as the fractional number in the range ΔV per unit range in V and given by the equation

$$f(V) = \frac{\Delta n_v}{n \Delta V} \quad (4-22)$$

Here, Δn_v is the number of cars in the speed range between V and $V + \Delta V$ as given by the best step type curve, and n is the total number of cars. The value of $f(V)$ is considered to be given for the mid-point of the range ΔV by Eq. (4-22). However, if ΔV is small enough, a smooth curve can be drawn between $f(V)$ and V as shown in Fig. 4-4b. This curve has essentially the same shape as the smoothed-out step type curve discussed above.

The resulting continuous function $f(V)$ is exceedingly useful, for the methods of the calculus may now be applied. For example, Eq. (4-22) can now be written in terms of differentials if n is sufficiently large, giving

$$f(V) = \frac{dn_v}{n dV} \quad (4-23)$$

In this equation, dn_v is the number of cars having speeds in the range from V to $V + dV$, or, more briefly, the number in dV .

To illustrate the meaning of $f(V)$, we rearrange Eq. (4-23) to yield the expression

$$f(V) dV = \frac{dn_v}{n} \quad (4-24)$$

Stated in words, $f(V) dV$ is equal to the fraction of the cars that are likely to have speeds in dV . An inspection of Fig. 4-4b will show that $f(V) dV$ is represented by the crosshatched area under the curve. This suggests that an integration of Eq. (4-24) may give additional information concerning the function $f(V)$. Upon integrating the right side of Eq. (4-24), we obtain n/n , or unity. Therefore, we have

$$\int_0^{\infty} f(V) dV = 1 \quad (4-25)$$

This simply means that the fraction of cars having speeds from 0 to ∞ is unity, for all the cars must be included in this range.

If the function $f(V)$ is known, then it will be easy to solve such problems as the following: How many cars had speeds over 65 m.p.h.? The result is obtained by solving Eq. (4-24) for dn_v , then summing the dn_v 's over the desired speed range. The result is obtained by integration, thus:

$$\text{Number having speeds over 65 m.p.h.} = n \int_{65}^{\infty} f(V) dV$$

The integration can be performed graphically by the use of Fig. 4-4b or analytically if the mathematical form of $f(V)$ is known.

4-9. Maxwell's Deduction of the Velocity Distribution Laws.—

For the purpose of the ensuing development,¹ we consider a fixed volume of gas at a constant temperature in which there are no convection currents or other disturbances of the equilibrium of the gas. The postulates fundamental to the molecular theory listed in Sec. 4-3 are considered to apply to the molecules of the gas. The last two postulates given in Sec. 4-3 are particularly important in what follows. These state, in effect, that the number of molecules concerned is very large and that the molecules have been colliding for a time sufficient to result in an equilibrium in the gas.

Besides the general assumptions mentioned above, Maxwell made the three basic assumptions that follow: (1) The velocities of the molecules are assumed to be distributed in a completely random fashion in space. This is equivalent to the statement that the distribution of velocity components of the molecules is

¹ V. ENGEL, A., and M. STEENBECK, "Elektrische Gasentladungen," Vol. I, p. 226, Verlag Julius Springer, Berlin, 1932.

the same along any three coordinate axes. (2) The velocity components along any three coordinate axes are assumed to be independent of one another. (3) The third assumption refers to the form of the law of distribution of molecular velocities and states that the fraction of the molecules contained in a certain range of speeds is a function only of the speed and that, moreover, this fraction is directly proportional to the range of speeds. This assumption also applies to the distribution of the *components* of the molecular velocities. The significance of these assumptions will become clearer as they are applied.

To apply Maxwell's postulates, we find it necessary to consider the mutually perpendicular components V_x , V_y , and V_z of the vector velocity $\mathbf{V}$ of a typical molecule as shown in Fig. 4-5a. Applying Maxwell's third postulate, we express the fraction of the molecules that are likely to have velocity components in the range from V_x to $V_x + dV_x$ (*i.e.*, in dV_x) as a function of the velocity component V_x , and directly proportional to the range dV_x , thus:

$$\frac{dn_{V_x}}{n} = f(V_x) dV_x \quad (4-26)$$

Here, n refers to the total number of molecules, and we recall that dn_{V_x} refers to the number of molecules in dV_x . We note that Eq. (4-26) has the same form as Eq. (4-24). Now, we apply Maxwell's first assumption when we consider that the fraction of the molecules which are likely to have velocity components in dV_y and dV_z will be given by the same form of function as in Eq. (4-26), namely, $f(V_y) dV_y$ and $f(V_z) dV_z$, respectively, in which the functions $f(V_y)$ and $f(V_z)$ will have exactly the same form as $f(V_x)$.

We next calculate the fraction of the molecules that *simultaneously* have velocity components in the range dV_x and dV_y . By Maxwell's second assumption, the molecules that have velocity components in dV_x will have y - and z -components of velocity distributed in the same fashion as those of all the molecules of the gas. Thus, of the molecules in dV_x , the fraction that has y -components in dV_y as well is equal to $f(V_y) dV_y$. Therefore, the fraction of all the molecules that have components in dV_x and dV_y will be the product of the two fractions, as follows:

$$\text{Fraction in } dV_x \text{ and } dV_y = f(V_x) dV_x f(V_y) dV_y \quad (4-27)$$

An extension of the argument to include the z -components will show that the fraction of the molecules that simultaneously have velocity components in dV_x , dV_y , and dV_z will be given by the following:

$$\text{Fraction in } dV_x dV_y dV_z = f(V_x)f(V_y)f(V_z) dV_x dV_y dV_z \quad (4-28)$$

To aid in understanding Eq. (4-28), we can imagine that the instantaneous velocities of all the molecules have been plotted as vectors setting out from the origin as in Fig. 4-5a. Such a

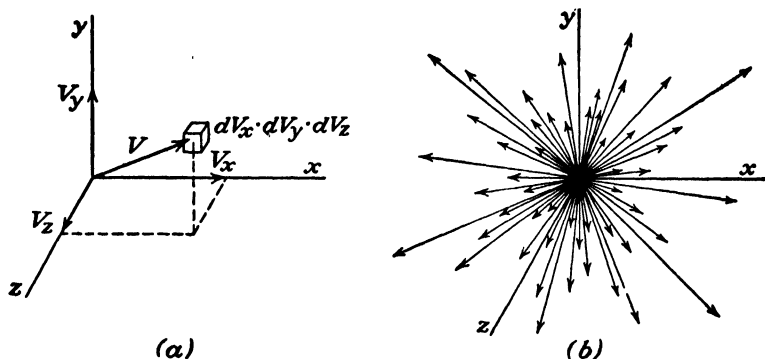


FIG. 4-5.—Velocity vectors.

representation is shown in Fig. 4-5b, although for an actual quantity of gas the number of molecules is so large that the space in Fig. 4-5b would be much more densely populated with vectors than as shown. This dense distribution of vectors is necessary if their distribution is to be accurately described by a continuous function, as in Eq. (4-28). The right side of Eq. (4-28) may now be interpreted as the fraction of all the velocity vectors whose tips fall in the elementary volume $dV_x dV_y dV_z$.

The fraction of the molecules whose vector tips end in the cube $dV_x dV_y dV_z$ can also be expressed as a function of the total velocity V of the molecules as $F(V) dV_x dV_y dV_z$; for, by assumption, the direction of V is immaterial. We also know that the total velocity is related to its components by the following:

$$V^2 = V_x^2 + V_y^2 + V_z^2 \quad (4-29)$$

For convenience, we replace the function $F(V)$ by a function of V^2 , letting $F(V) = g(V^2)$. Using this function, we now equate the expression for the fraction of the molecules in $dV_x dV_y dV_z$

in terms of the total velocity to the expression in terms of the components of velocity [Eq. (4-28)] and obtain the equation

$$f(V_x)f(V_y)f(V_z) dV_x dV_y dV_z = g(V^2) dV_x dV_y dV_z \quad (4-30)$$

Therefore,

$$f(V_x)f(V_y)f(V_z) = g(V_x^2 + V_y^2 + V_z^2) \quad (4-31)$$

The form of Eq. (4-31) suggested to Maxwell that the following functions would constitute a possible solution:

$$\left. \begin{aligned} f(V_x) &= Ae^{BV_x^2} \\ f(V_y) &= Ae^{BV_y^2} \\ f(V_z) &= Ae^{BV_z^2} \end{aligned} \right\} \quad (4-32)$$

A substitution of these functions into Eq. (4-31) will show that they satisfy this equation. The solution of Eq. (4-31) as given in Eqs. (4-32) can also be obtained by direct mathematical means.¹ The constants A and B in Eqs. (4-32) must be chosen to suit the physical conditions of the problem. It is obvious that A must be a positive number, for a negative fractional number of molecules has no meaning. One might think that B could be either positive or negative; but a positive B would lead to a distribution that increases indefinitely as V_x increases, and this cannot be true. Therefore, B is a negative number.

To evaluate the constants A and B , we again consider the velocity vector pictures in Fig. 4-5. The constant A can be eliminated in terms of the total number of molecules n . To accomplish this, first calculate the number $dn_{\square}$ of molecules whose vector tips lie in the cube $dV_x dV_y dV_z$ by multiplying the fraction of the molecules given by Eq. (4-28) by the total number of molecules, obtaining

$$dn_{\square} = n f(V_x)f(V_y)f(V_z) dV_x dV_y dV_z = n g(V^2) dV_x dV_y dV_z \quad (4-33)$$

But, by Eqs. (4-32), this reduces to the following:

$$dn_{\square} = A^3 n e^{B(V_x^2 + V_y^2 + V_z^2)} dV_x dV_y dV_z = A^3 n e^{BV^2} dV_x dV_y dV_z \quad (4-34)$$

Now the concentration of the vector tips, *i.e.*, the number per unit volume, at a distance V from the origin, as in Fig. 4-5b, is given by

$$\frac{dn_{\square}}{dV_x dV_y dV_z} = A^3 n e^{BV^2} \quad (4-35)$$

¹ General reference 4-3, p. 78.

To complete the calculation of the total number of vectors, we consider a spherical shell of radius V and of radial thickness dV centered at the origin. The number dn_v in this spherical shell will be equal to the number per unit volume, given by Eq. (4-35), multiplied by the volume of the shell, which is $4\pi V^2 dV$. Therefore, the number of vector tips in the spherical shell is as follows:

$$dn_v = 4\pi A^3 n V^2 e^{BV^2} dV \quad (4-36)$$

We note that dn_v also is the number of molecules that have speeds between V and $V + dV$. If Eq. (4-36) is integrated throughout the space in Fig. 4-5b, the value of the integral will be the total number of molecules n . Thus the integral of the left side of the Eq. (4-36) will be n , and therefore

$$n = 4\pi A^3 n \int_0^\infty V^2 e^{BV^2} dV \quad (4-37)$$

The integral will give a finite value only if B is taken as negative, confirming our previous conclusion concerning the sign of B . Since the exponent of e must be a pure number, the dimensions of the constant B must be $1/(\text{velocity})^2$. Therefore, let

$$B = -\frac{1}{c_0^2} \quad (4-38)$$

Here c_0 is a parameter that has the dimensions of a velocity and is a constant for the gas in the given condition but that can vary with the condition of the gas. Now Eq. (4-37) reduces to the following:

$$1 = 4\pi A^3 \int_0^\infty V^2 e^{-\frac{V^2}{c_0^2}} dV \quad (4-39)$$

The value of the definite integral¹ is $c_0^3 \sqrt{\pi/4}$, so the constant A is found to have the value

$$A = \frac{1}{\sqrt{\pi} c_0} \quad (4-40)$$

By putting this value of A in Eq. (4-32) and remembering Eq. (4-26), we can write the final form for the distribution of the velocity components as follows:

¹ PIERCE, B. O., "A Short Table of Integrals," 2d ed., Formula (494), Ginn and Company, Boston, 1910.

$$dn_{V_x} = n f(V_x) dV_x = \frac{n}{\sqrt{\pi} c_0} e^{-\frac{V_x^2}{c_0^2}} dV_x \quad (4-41)$$

Also, from Eq. (4-36), we obtain the final form of the distribution of speeds irrespective of direction.

$$dn_V = n f(V) dV = \frac{4nV^2}{\sqrt{\pi} c_0^3} e^{-\frac{V^2}{c_0^2}} dV \quad (4-42)$$

Equations (4-41) and (4-42) are often called Maxwell's distribution laws or the Maxwell-Boltzmann distribution laws.

A brief consideration of the form of the function in Eq. (4-42) will reveal the meaning of the constant c_0 . We note that the speed V appears in the two factors, namely, V^2 and $e^{-\frac{V^2}{c_0^2}}$. At zero speed, *i.e.*, $V = 0$, the value of the function is zero. For increasing values of V , the V^2 factor increases, and the exponential factor decreases. For very large values of V compared with c_0 , the exponential term approaches zero so rapidly that the function also approaches zero. In the intermediate speed range the function remains positive and therefore goes through a maximum value. Upon differentiating $f(V)$ with respect to V and equating to zero, we find that the maximum of the function occurs when V equals c_0 . Therefore, the speed c_0 may be defined to be the *most probable* speed of the molecules. In Sec. 4-11, we shall show that c_0 depends only on the temperature of the gas and on the kind of gas.

4-10. Representation of Maxwell's Distribution Laws.—The distribution of speeds of the molecules of a gas is in some respects similar to the distribution of speeds of cars discussed in Sec. 4-8. However, in dealing with distributions of velocities of molecules, we are concerned not only with the distribution of speeds irrespective of direction in space, but also with the distribution of velocity components along a particular axis. In the following discussion, we consider the distribution of speeds first.

Let us examine the mathematical function expressing the distribution of speeds given in the following equation, which follows directly from Eq. (4-42):

$$\frac{dn_V}{n} = \frac{4V^2}{\sqrt{\pi} c_0^3} e^{-\frac{V^2}{c_0^2}} dV \quad (4-42a)$$

We recall that dn_v is the number of molecules having speeds between V and $V + dV$, and n is the total number of molecules; the ratio of dn_v to n is then the fraction of the total number of molecules having speeds in the range dV , and c_0 is the *most probable speed* of the molecules. The distribution given by Eq. (4-42a) could be represented by a graph similar to that in Fig. 4-6b if a value were assigned to c_0 . However, this curve would apply only to one particular condition of the gas. Since it is desirable to represent the distribution in a sufficiently

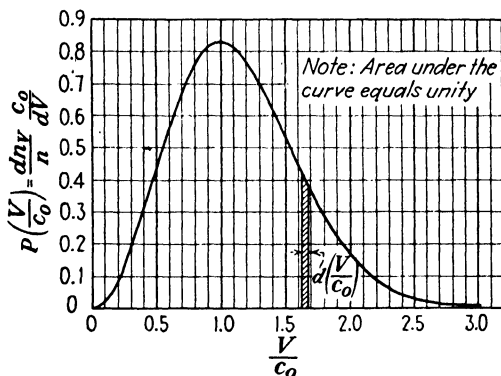


FIG. 4-6.—Distribution curve for molecular speeds (Maxwell-Boltzmann distribution).

general way so that a single curve can be applied to the gas under any set of conditions, we proceed by rearranging Eq. (4-42a) to give a function of V/c_0 as follows:

$$P\left(\frac{V}{c_0}\right) = \frac{dn_v}{n} \frac{c_0}{dV} = \frac{4}{\sqrt{\pi}} \left(\frac{V}{c_0}\right)^2 e^{-\left(\frac{V}{c_0}\right)^2} \quad (4-43)$$

We note that this function is the fractional number of molecules in the range dV per unit range in V/c_0 . If we plot the value of $P(V/c_0)$ as ordinate against the ratio V/c_0 as abscissa, the resulting curve is shown in Fig. 4-6. The rising nature of the curve for small values of the abscissas may be attributed to the predominating influence of the $(V/c_0)^2$ factor. The exponential factor becomes increasingly important as V/c_0 increases, and finally causes the curve to approach zero very rapidly as V/c_0 exceeds a value of about 3.

An insight into the meaning of the curve in Fig. 4-6 can be gained by inquiring into the interpretation to be put upon the

area under the curve within a range of abscissas equal to the differential increment $d(V/c_0)$, i.e., dV/c_0 . This area, shown crosshatched in Fig. 4-6, is equal to dV/c_0 multiplied by the ordinate $dn_V c_0/n dV$; and the result is dn_V/n . Therefore, the area under the curve in the range $d(V/c_0)$ is equal to the fraction of the molecules that have speeds in the range dV . Furthermore, the total area under the curve will represent the summation of fractions of the type dn_V/n over all the speed ranges, thus including all the molecules. Therefore, the total area under the curve will equal n/n , or unity. Consequently, the fraction of the molecules that have speeds in an extended range, say from V_1 to V_2 , will equal the area under the distribution curve between the limits V_1/c_0 and V_2/c_0 .

Another aspect of the distribution curve will be made clear if we consider a fixed range of speeds ΔV , but at different values of the speed V , for then we see that the height of the curve is proportional to the fraction of the molecules in ΔV at the different speeds. A range ΔV taken at a speed V equal to c_0 , i.e., $V/c_0 = 1$, contains a greater fraction of the molecules than any other equal speed range. This shows the reason for calling c_0 the most probable speed.

Some numerical calculations will be useful to illustrate the quantities involved in Eq. (4-43) and in Fig. 4-6. Consider 1 cu. cm. of gas at standard conditions, which therefore contains 2.69×10^{19} molecules, and assume that the kind of gas is such that the most probable speed c_0 is 500 m./sec. Suppose we desire to know the number of molecules that, at any instant, are likely to have speeds between 500.0 and 500.1 m./sec. From the curve in Fig. 4-6 at $V/c_0 = 1$, we note that the value of the ordinate is 0.83. To find Δn_V , this ordinate value must be multiplied by n and by ΔV and divided by c_0 . The resulting value of Δn_V is 44.7×10^{14} . If the calculation is repeated at $V = 1,000$, i.e., at twice the most probable speed, and if ΔV is still 0.1 m./sec., we find that $\Delta n_V = 8.91 \times 10^{14}$, or 20 per cent of the number in an equal range at $V/c_0 = 1$. When the velocity is three times the most probable speed, we find that Δn_V is 0.3 per cent of the value at the most probable speed. These results show how rapidly the function decreases as the speed V increases.

A general representation of the distribution of the *components of the velocities* is obtained by analyzing Eq. (4-41a) which comes directly from Eq. (4-41).

$$\frac{dn_{V_x}}{n} = \frac{1}{c_0 \sqrt{\pi}} e^{-\frac{V_x^2}{c_0^2}} dV_x \quad (4-41a)$$

Here dn_{v_x} is the number of molecules that have x -components of velocity between V_x and $V_x + dV_x$. Thus, upon rearranging Eq. (4-41a), we may write

$$p\left(\frac{V_x}{c_0}\right) = \frac{dn_{v_x}}{n} \frac{c_0}{dV_x} = \frac{1}{\sqrt{\pi}} e^{-\frac{V_x^2}{c_0^2}} \quad (4-44)$$

Expressions identical to Eq. (4-44) except for a change in the subscripts give the distributions of the y - and z -components of velocities. The right side of Eq. (4-44) is a function of the ratio V_x/c_0 , and so this ratio has been used for the abscissa, and the value of the function, $p(V_x/c_0)$, for the ordinate in the general

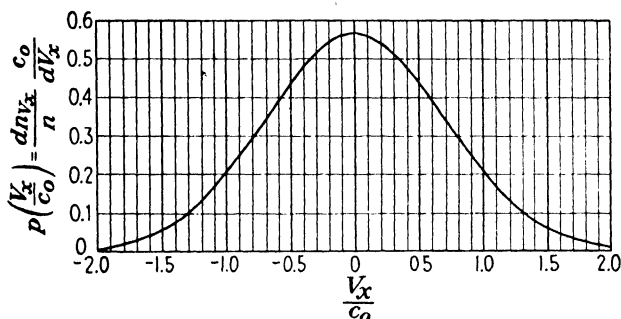


FIG. 4-7.—Distribution curve for the components of molecular velocities.

distribution curve shown in Fig. 4-7. This curve is symmetrical around the ordinate axis, for positive and negative components of velocity occur in equal numbers. Furthermore, the most probable velocity component is zero, as shown by the maximum ordinate of the curve.

If we refer to Fig. 4-5b, we can visualize the number of molecules having components in dV_x by imagining a differential slab dV_x thick with its faces perpendicular to the x -axis. Then, dn_{v_x} will be the number of vectors whose tips end in the differential slab, for all these vectors will have x -components in the range dV_x . If this slab is taken at the origin, it will include more vector tips than if it is taken at any other value of V_x .

Direct tests of the distribution of the velocities of the molecules have been obtained¹ in connection with experiments on molecular beams, and the results agree satisfactorily with the theory. Other experiments show¹ that a Maxwellian distribution of

¹ General reference 4-3, pp. 113-138.

velocities is found among the electrons thermionically emitted from a hot body. Tests with probe electrodes (see Sec. 10-6) show that this distribution is also found among the electrons in some of the regions in a low-pressure discharge or glow discharge, for example, in the Faraday dark space and in the luminous positive column. Where the electric field is strong, there are departures from the Maxwellian distribution, although in some instances the experimental results show that the electrons are distributed according to two superimposed Maxwellian distributions.

The knowledge we have gained of the distribution of the components of velocity will become useful when we consider the numbers of particles that collide with a surface. Another application will arise when we calculate the rate of escape of electrons from heated filaments, *i.e.*, in connection with the theory of thermionic emission of electrons.

4-11. The Root-mean-square and Arithmetical Average Speeds.—The distribution of the speeds of the molecules having been determined, it is important to inquire as to the meaning of the average speed C used in the simple derivation for the pressure given in Sec. 4-4. Since the pressure exerted by a gas on the walls of its container is simply the average force per unit area exerted on the walls by the impacts of molecules having different speeds, it is not difficult to see¹ that the speed C^2 in Eq. (4-7) is the average of the squares of the speeds of the molecules. In short, C is the *root-mean-square speed* (r.m.s. speed). To calculate the r.m.s. speed on the basis of the Maxwellian speed distribution, we multiply the square of the speed V by the number of molecules having speeds in dV , integrate so as to include the total number of molecules, divide by the total number of molecules, and take the square root, as follows:²

$$C = \sqrt{\frac{\int_{V=0}^{V=\infty} V^2 dn_v}{n}} \quad (4-45)$$

¹ See general reference 4-1, pp. 8-10.

² This formula is analogous to the expression for the r.m.s. value of an alternating current of period T_1 , namely, $I_{r.m.s.} = \sqrt{\int_0^{T_1} i^2 dt / T_1}$. Although the i^2 term in this expression is weighted in the averaging process by the factor dt ; in Eq. (4-45) the V^2 term is weighted by the number dn_v , which is itself a function of V .

We substitute for dn_v its value given by Eq. (4-42) and obtain

$$C = \sqrt{\frac{4c_0}{\sqrt{\pi}}} \int_0^{\infty} \frac{V^4}{c_0^4} e^{-\frac{V^2}{c_0^2}} dV$$

If we let $V/c_0 = w$, the expression for C reduces to the following:

$$C = \sqrt{\frac{4c_0^2}{\sqrt{\pi}}} \int_0^{\infty} w^4 e^{-w^2} dw$$

The definite integral has the value¹ $\frac{3}{8} \sqrt{\pi}$; therefore,

$$C = \sqrt{\frac{3}{2}} c_0 = 1.225 c_0 \quad (4-46)$$

This shows that the r.m.s. speed is 22.5 per cent higher than the most probable speed.

The *arithmetical average speed*, designated by the symbol $\bar{c}$, is useful in some types of calculations. It is found by multiplying the speed V by the number of molecules in dV , integrating over all the molecules, and then dividing by the total number of molecules, as follows:

$$\bar{c} = \frac{\int_{v=0}^{v=\infty} V dn_v}{n} \quad (4-47)$$

In effect the integration consists in adding the speeds of all the molecules; hence, dividing by the total number gives the arithmetical average speed. Upon substituting for dn_v from Eq. (4-42), we find

$$\bar{c} = \frac{4}{\sqrt{\pi}} \int_0^{\infty} \frac{V^3}{c_0^3} e^{-\frac{V^2}{c_0^2}} dV$$

If we again let $V/c_0 = w$, this equation reduces to

$$\bar{c} = \frac{4c_0}{\sqrt{\pi}} \int_0^{\infty} w^3 e^{-w^2} dw$$

The value² of the definite integral is $\frac{1}{2}$; therefore, we obtain the result

$$\bar{c} = \frac{2}{\sqrt{\pi}} c_0 = 1.129 c_0 \quad (4-48)$$

¹ General reference 4-3, p. 654. Also B. O. Pierce, "Short Table of Integrals," 2d ed., Formula (494), Ginn and Company, Boston, 1910.

² General reference 4-3, p. 654. Also, B. O. Pierce, *op. cit.*, Formula (481).

This shows that the arithmetical average speed is greater than the most probable speed but less than the r.m.s. speed. From Eqs. (4-46) and (4-48), we obtain the relation

$$\bar{c} = \sqrt{\frac{8}{3\pi}} C = 0.92C \quad (4-49)$$

Since the speed C of Secs. 4-4 and 4-6 has been identified as the r.m.s. speed, we can derive a useful formula relating the most

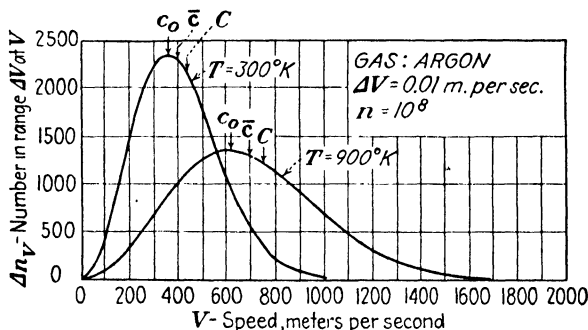


FIG. 4-8.—Distribution of speeds for argon.

probable speed to the temperature and the kind of gas. We had the relation

$$C = \sqrt{\frac{3kT}{m}} \quad (4-17)$$

We recall that k is the Boltzmann constant and m is the mass of a single molecule. Upon eliminating C between Eqs. (4-46) and (4-17), we find :

$$c_0 = \sqrt{\frac{2kT}{m}} \quad (4-50)$$

This last expression is important because the value of c_0 must be known before the velocity distribution laws can be applied to a numerical problem. Thus, Eq. (4-50) is needed in addition to Eqs. (4-41) and (4-42) to describe completely the distribution of the velocities of the molecules of a gas at any given temperature.

The curves shown in Fig. 4-8 have been drawn to illustrate the values of the various average speeds in relation to the distribution of speeds. These curves for argon gas at two different temperatures give the value of Δn_V as a function of the speed V for the chosen value of ΔV of 0.01 m./sec., the value of n having been

taken as 100 million (10^8). The curves were obtained by calculating c_0 for one of the given temperatures and then obtaining Δn_v from Fig. 4-6 by the method employed in Sec. 4-10, repeating the process for the other temperature. Figure 4-8 not only shows the relative values of the most probable, arithmetical average, and r.m.s. speeds at the two temperatures but also shows that as the temperature increases the height of the curve decreases. This is to be expected for this type of distribution curve, since the same number of molecules is distributed through a wider speed range than before.

4-12. Free Paths of Molecules and Electrons.—In the discussions of molecular motions in the preceding sections the molecules were thought of as small solid spheres which move in straight-line paths through space during the intervals between elastic impacts with other molecules or the walls. The actual molecules are known to have a much more complex structure than we have assumed thus far. The modern view is that an atom consists of a massive, positively charged nucleus surrounded by a sufficient number of electrons so that the atom as a whole is electrically neutral. The electrons are conceived to move in circular or elliptical paths about the nucleus. The nucleus and the electrons are each of relatively small dimensions compared with the atomic diameter. Experimental evidence and theoretical calculations indicate the diameters of the nucleus and the electrons to be of the order of 10^{-13} cm., and the "diameter" of the atom is roughly 100,000 times as great, or of the order of 10^{-8} cm. Thus, the volume occupied by the atom consists mainly of free space, and questions regarding the shape and size of atoms or molecules cannot be answered very definitely without a knowledge of the positions or paths of the electrons and nuclei that make up the molecule. When two neutral molecules approach one another within one or two molecular diameters, their systems of charges experience repulsive forces, probably electrostatic in nature, which increase very rapidly as the distance between the two molecules decreases. Thus a "collision" between two molecules is actually a deflection of their paths caused by electrical forces acting between them, and the terms "collision" and "molecular diameter" have no precise meanings without careful definition. For the purposes of the following discussion, however, we retain the simpler picture of the perfectly

elastic spherical molecule and adhere to the assumptions listed in Sec. 4-3. The relations obtained by considering this idealized molecule must be regarded as approximate for the actual molecules.

The distance that a molecule moves between successive impacts is called a *free path*. The preceding discussions of the motions of the molecules lead us to conclude that a molecule may have a free path that is very short or even zero, such as is experienced by molecule *B* in Fig. 4-3 after its impact with molecule *A*; or it may have a free path as long as, say 100,000 molecular diameters. If a molecule were singled out and the projection of its path on a plane were mapped for a period of time involving several collisions, the resulting figure might appear as in Fig. 4-9.

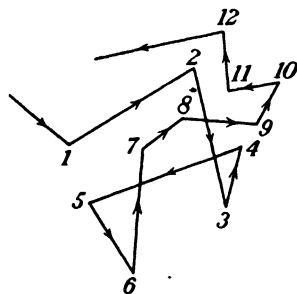


FIG. 4-9.—Path of a molecule.

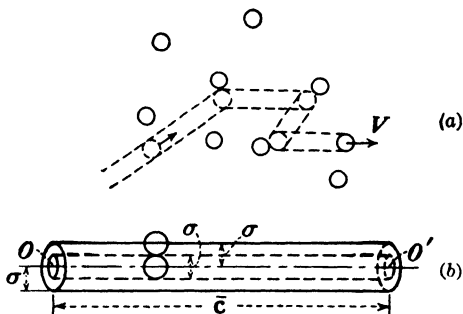


FIG. 4-10.

This figure also illustrates the wide variation in the lengths of the free paths that is possible. The *average* length of the free paths, averaged over a large number of paths, is called the *mean free path* of the molecule. It is to be expected that the mean free path will be less for large molecules than for small molecules and that it will be less as the molecular concentration increases, for both factors increase the likelihood of collisions.

An estimate of the effect of these two factors on the mean free path can easily be made. Assume that all but one of the molecules are fixed in space and that the moving molecule has the speed $\bar{c}$. This molecule will travel a zigzag path between impacts as seen in Fig. 4-10a. In Fig. 4-10b the zigzag path of the center of the moving molecule has been straightened out and is represented by the line OO' , which is taken as $\bar{c}$ m. in length, i.e., the distance traveled in 1 sec. It is proposed to calculate the

probable number of collisions experienced by the molecule in traveling the distance $\bar{c}$ m. by calculating the average number of molecules that would be near enough to the path of the moving molecule to cause a collision. From Fig. 4-10*b*, it is clear that the molecules whose centers are closer than σ to the line OO' would be struck by the moving molecule, where σ designates the diameter of the molecules. All the molecules whose centers lie within the outer circular cylinder shown in Fig. 4-10*b* would thus cause collisions. The number of molecules in this cylinder is the molecular concentration N multiplied by the volume of the cylinder, which is $\pi\sigma^2\bar{c}$; the result is $\pi\sigma^2N\bar{c}$. The average distance between impacts (mean free path) will be the distance traveled $\bar{c}$, divided by the number of impacts, which was just computed to be $\pi\sigma^2N\bar{c}$. Designating the mean free path by the symbol λ , we then have

$$\lambda = \frac{\bar{c}}{\pi\sigma^2N\bar{c}} \quad \text{or} \quad \lambda = \frac{1}{\pi\sigma^2N}$$

This simple derivation has given the correct variation of the mean free path with molecular concentration and molecular diameter, but the assumption of stationary molecules leads to a numerical coefficient that is incorrect for an actual gas in which the molecules are in motion.

When the molecules are moving, a given molecule has more impacts per second than the number calculated above. This is shown by the following value of the number of collisions per second, Z , calculated¹ on the basis of the Maxwellian distribution of velocities:

$$Z = \sqrt{2}\pi\sigma^2N\bar{c} \quad (4-51)$$

This quantity is often called the *collision frequency*. Since the mean free path is the average velocity $\bar{c}$ divided by the collision frequency Z , we find

$$\lambda = \frac{1}{\sqrt{2}\pi\sigma^2N} \quad (4-52)$$

This equation states that the mean free path is inversely proportional to the cross-sectional area of the molecule and also inversely proportional to the molecular concentration.

If the effective diameter of a molecule is known, its mean free path may be calculated by Eq. (4-52); or, conversely, if the mean

¹ General reference 4-2, p. 37.

free path is known, its diameter may be calculated. To find the order of magnitude of the mean free path of molecules under standard conditions, let us assume a molecular diameter of 3×10^{-10} m. and calculate λ by means of Eq. (4-52). We find as a result that $\lambda = 0.925 \times 10^{-7}$ m. This is a very short distance; but when it is compared with the molecular diameter, we find that λ is over 300 times the molecular diameter. The mean free path can also be compared with the average distance between the molecules, as follows: Each molecule may be considered to occupy, on the average, a volume of $1/N$ cu. m. which can be represented by a cube having an edge $1/N^{1/3}$ m. in length. This distance may also be taken to be the approximate average distance between molecules. Under standard conditions, this distance works out to be 3.34×10^{-9} m., and therefore the mean free path of the molecule considered above is nearly 28 times the average distance between the molecules. This shows that the two free paths illustrated in Fig. 4-10a are relatively short compared with the particular mean free path under discussion.

If the molecular concentration in Eq. (4-52) is replaced by its value given by Eq. (4-12), we obtain an expression that shows how changes in temperature and pressure affect the mean free path, namely,

$$\lambda = \frac{RT}{\sqrt{2} \pi \sigma^2 N_0 p} \quad (4-53)$$

Thus, the mean free path is proportional to the absolute temperature if the pressure remains constant and inversely proportional to the pressure if the temperature remains constant. For example, if the mean free path of the molecule considered in the foregoing paragraph is to be calculated at the low pressure of, say 10^{-8} atm. (7.6×10^{-6} mm. Hg), the result would be 10^8 times the previous value, or 9.25 m. Under high-vacuum conditions, therefore, the mean free path of the particles of the gas may greatly exceed the dimensions of the apparatus that contains the gas. Collisions with walls and electrodes will thus be much more frequent than collisions between molecules.

The numerical value of the mean free path of a gas is usually obtained from experimental tests of some physical property of the gas that depends on the mean free path of its molecules. For example, the viscosity coefficient η of a gas calculated on the

basis of the motions of the molecules is given by the formula¹

$$\eta = b\rho\bar{c}\lambda \quad (4-54)$$

Here b is a numerical coefficient that is known from the theory of the viscosity on the basis of which Eq. (4-54) was derived and ρ is the density of the gas. Experimental measurements of η , therefore, yield values of the mean free path λ , for the density ρ and the mean velocity $\bar{c}$ are known. The values of the mean free paths for various gases listed in Appendix III were obtained in this manner. These values are used as a basis for computing the mean free path under pressures and temperatures other than the reference values given in the table. A convenient formula for this purpose may be derived by letting λ_0 , p_0 , and T_0 be the mean free path, pressure, and temperature under one set of conditions and λ_1 , p_1 , and T_1 be the corresponding values under another set of conditions, and then calculating the ratio λ_1/λ_0 by means of Eq. (4-53), thus obtaining

$$\frac{\lambda_1}{\lambda_0} = \frac{T_1 p_0}{T_0 p_1} \quad (4-55)$$

When dealing with gaseous discharges, we shall be concerned with the free paths of the free electrons. An electron can approach more closely to a neutral molecule, before the repulsive forces between the two are large enough to cause deviations in their paths, than a molecule can. Some experiments even indicate that high-speed electrons can pass through a molecule. However, we shall assume that an electron is deflected from a molecule as if a collision occurred between elastic spheres and shall assume that the electron radius is negligible compared with the radius of the molecule. To calculate the mean free path of an electron immersed in a gas, we imagine that a single electron is projected into a gas in which the molecules are stationary, just as in the preliminary calculation of the mean free path of a molecule. Since the center of the electron is considered to be deflected from the surface of the molecules, we see that the outer cylinder in Fig. 4-10*b* would now shrink to the same size as the inner cylinder; *i.e.*, it will have a radius of only $\sigma/2$. Following the same line of reasoning as before, we find that the mean free

¹ General reference 4-3, pp. 207, 214-216.

path of the electron λ_e calculated on this basis is given by the expression

$$\lambda_e = \frac{4}{\pi \sigma^2 N} \quad (4-56)$$

This applies to the condition where a small percentage of free electrons is mixed with a gas of molecular diameter σ and of concentration N . In the derivation for the mean free path of the gas molecules a factor of $1/\sqrt{2}$ was introduced to take account of the motions of the molecules. As we have already seen (Sec. 4-7), the thermal velocities of electrons in equilibrium with gas molecules are relatively high. As a consequence, Eq. (4-56) needs no correction for the motions of the molecules.

By comparing Eqs. (4-56) and (4-52), we see that

$$\lambda_e = 4 \sqrt{2} \lambda \quad (4-57)$$

That is, the mean free path of an electron is $4 \sqrt{2}$ times the mean free path of the molecules of the gas in which the electron is immersed.

An illustration of the application of Eq. (4-57) occurs in the solution of the following problem: A high-vacuum cathode-ray tube is 0.3 m. long from cathode to screen, and it is desired that the mean free path of the electrons in the tube shall be at least 3 m. so that only a small number of electrons will have collisions with gas molecules. What pressure is permissible in the tube, if the residual gas is assumed to be nitrogen? The permissible mean free path of the nitrogen molecules is, by Eq. (4-57), equal to 0.53 m. From the table in Appendix III, we find that λ for nitrogen at $p_0 = 1$ atm. and $T_0 = 273^\circ\text{K.}$ is 5.96×10^{-8} m. If we assume that the cathode-ray tube works at a temperature of 300°K. , we can use Eq. (4-55) to find that the pressure p should be no higher than 0.124×10^{-6} atm. (0.94×10^{-4} mm. Hg).

4-13. Law of Distribution of Free Paths.—A knowledge of the mean free path of the ions in a discharge is often helpful toward gaining an understanding of the behavior of the discharge, but a more detailed study of the lengths of the free paths is sometimes necessary*. The picture of the chaotic motions of the molecules discussed in connection with the distribution of the speeds of the molecules suggests that there will also be a wide range of distribution of the free-path lengths. The length of a free path is entirely dependent on the chance that the molecule's free flight will be interrupted by a collision. The law of distribution of the free paths will now be derived by applying the

laws of chance to the free-path lengths of a large number of molecules.

Let us consider a large number n_0 of molecules each of which is setting out on a free path just after having had a collision. Let the distance along each path be represented by x , and let n represent the number of molecules that *have not yet had a collision* at the end of a path of length x . That is, n is the number of molecules whose free paths exceed the distance x . Furthermore, let dn be the increase in n in the incremental distance dx . We realize that dn will be numerically equal to the number of molecules that have collisions in the distance dx but that it will be a negative quantity because n is decreasing. Under conditions of random motions of the molecules the number dn should be proportional to n , because a collision is equally likely for each of the molecules and n have not experienced a collision in the distance dx . The number dn should also be proportional to the distance dx ; therefore, we can write

$$dn = -cn \, dx \quad (4-58)$$

Here c is a proportionality constant whose value is still to be determined. We rearrange Eq. (4-58) to obtain the following form

$$\frac{dn}{n} = -c \, dx \quad (4-59)$$

Upon integrating this relation, we obtain the result that

$$n = ke^{-cx} \quad (4-60)$$

In this relation, k is an integration constant that must be determined from the physical conditions of the problem. This is accomplished by considering that, at $x = 0$, $n = n_0$, the total number of molecules being considered; therefore, $k = n_0$, and Eq. (4-60) reduces to

$$n = n_0 e^{-cx} \quad (4-61)$$

The constant c may be calculated in terms of the mean free path λ as follows: We formulate the expression for the mean free path as the sum of all the free-path lengths divided by the total number of paths n_0 , obtaining the expression

$$\lambda = \frac{\int_{x=0}^{x=\infty} x(-dn)}{n_0} \quad (4-62)$$

In this expression, $-dn$ is a positive quantity because of the way n was previously defined. From Eq. (4-61), we have $dn = -cn_0e^{-cx} dx$; therefore,

$$\lambda = \int_0^{\infty} cxe^{-cx} dx$$

Upon evaluating the integral, we find that

$$\lambda = \frac{1}{c} \quad (4-63)$$

Consequently, the constant c in Eq. (4-61) can be replaced by $1/\lambda$, and we have

$$n = n_0 e^{-\frac{x}{\lambda}} \quad (4-64)$$

This equation states that the number of free paths that exceed a distance x varies as an exponential function of x . The ratio of n to n_0 may be plotted as a function of the ratio x/λ , as seen in Fig. 4-11.

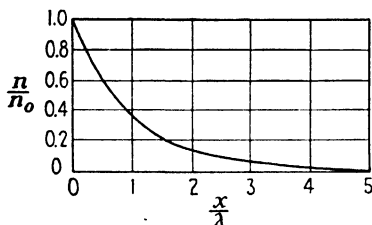


FIG. 4-11.—The fraction of the free paths that exceed the distance x , in terms of x/λ .

Equation (4-64) and Fig. 4-11 show that only 36.8 per cent of the free paths exceed the mean free path. Furthermore, the percentage drops rapidly as longer paths are considered; for example, only 0.67 per cent

of the paths exceed five times the mean free path.

If we are interested in the number of free paths that end in the range dx , we make use of the known variation of the number n of free paths that exceed the distance x with x [Eq. (4-64)], as follows: Since n is a decreasing function of x , dn/dx is negative, and we must take the absolute value of the differential of Eq. (4-64) to find the number of free paths that end in dx ; we thus find

$$|dn| = \frac{n_0}{\lambda} e^{-\frac{x}{\lambda}} dx \quad (4-65)$$

If Eq. (4-65) is rearranged to give the usual form for a distribution function [see Eq. (4-24)], we should write

$$\frac{|dn|}{n_0} = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx \quad (4-66)$$

This equation, therefore, expresses the fractional number of free paths that have lengths in the range dx .

Although the deductions concerning the distribution of free paths have been based on the motions of a large number of molecules, the results would be expected to apply to the distribution of the free paths of a single molecule, provided that a large number of successive free paths are considered. This follows from the fact that all the molecules are identical; and if enough free paths are considered, it makes no difference whether these free paths are the free paths of one molecule or of many molecules.

The development of means of generating beams of electrons or of ions, such as were discussed in Chap. III, has made possible a direct experimental observation of the free paths of these particles. In general, the method consists in projecting the beam of ions into a chamber containing gas at a reduced pressure and then testing for the decrease in the number of ions remaining in the beam along the length of the beam. The ions are deflected out of the beam at the end of their free paths, and the number

of ions per unit length remaining in the beam should decrease as an exponential function of the distance, in accordance with the results of the analysis at the beginning of this section. This expectation is fulfilled in the experiments. However, the rate of decrease of the number of ions in the beam is observed to depend on the velocities of the ions. In other words, the mean free path is observed to depend on the velocities of the ions. The relative magnitude of the variation in the mean free paths obtained by this method is shown in Fig. 4-12, which shows how the mean free paths of lithium, potassium, and cesium ions moving in neon gas depend on the initial energy of the ions in the beam. These curves have been adapted from the results of Ramsauer and Beeck.¹ The short lines at the right of the figure give the so-called kinetic-theory values of the mean free

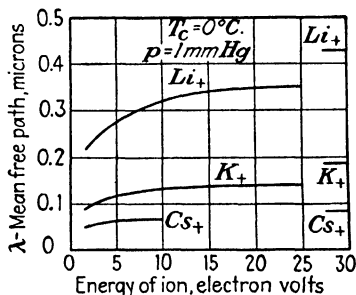


FIG. 4-12.—Mean free paths of alkali ions projected into neon.

¹ *Ann. Physik*, **87**, 18, 1928.

paths, *i.e.*, the values based on viscosity measurements. As seen from the figure, there is fair agreement between the kinetic-theory values and the experimental values for the higher speeds of the ions, but the agreement is poorer for the lower speeds.

Similar results for the experimental values of the mean free paths of electrons, together with the kinetic-theory values, are shown in Fig. 4-13. These curves show that the slow electrons

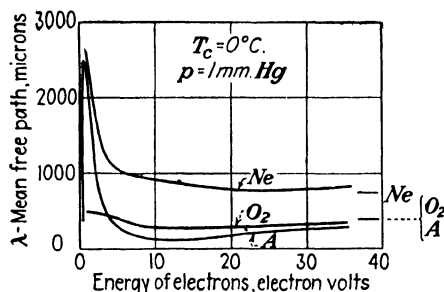


FIG. 4-13.—Mean free paths of electrons in various gases. (Based on data of Ramsauer; Ramsauer and Kollath; and Brüche. See Landolt-Bornstein Tabellen, 5 Auflage, Zweiter Erg., Erster Teil, pp. 70–72.)

have abnormally long path lengths in the inert gases but that the faster electrons behave more nearly according to the expectations of the theory.

The differences between theory and experiment shown in Figs. 4-12 and 4-13 will not be wondered at if one considers the

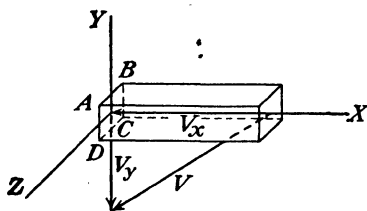


FIG. 4-14.

complex nature of the actual collision process. We must therefore consider the theory as only an approximate one, but one that is nevertheless very useful as an aid in visualizing the conditions in discharges and in making *order-of-magnitude* calculations dealing with

the motions of ions and electrons in discharges.

4-14. Calculation of the Number of Molecules Crossing a Plane Surface.¹—Though we have considered the energies, speeds, and free paths of the molecules, we have not yet considered how many molecules cross a plane surface in the gas in unit time. This information is useful in the study of evaporation

¹ General reference 4-4, p. 17.

and adsorption phenomena and will be used in the calculation of currents collected by probes in highly ionized regions.

Let us calculate the number of molecules that strike unit area $ABCD$ taken in the yz -plane as shown in Fig. 4-14. We may imagine that this area is located anywhere in the gas space; then we specify that we shall count only the molecules moving from right to left through the surface and disregard those moving in the opposite direction. Under equilibrium conditions, there will be equal numbers crossing the surface in both directions. If the area $ABCD$ is thought of as a portion of the boundary of the gas, then the number to be calculated will represent the number of molecules striking a unit area of the boundary per second.

If we resolve the velocity vector V of a single molecule into its x -, y -, and z -components, we see that it is only the x -component which will bring the molecule toward the surface $ABCD$ in the yz -plane. To take account of the distribution of the velocities, we consider the molecules having x -components in the range from V_x to $V_x + dV_x$. Now we see that all the molecules in the cylinder V_x long erected on $ABCD$ as a base which have velocities in the range V_x to $V_x + dV_x$ will cross the yz -plane in 1 sec. For example, a molecule starting from the right end of the cylinder with a velocity V_x will just reach the surface $ABCD$ at the end of 1 sec., on the assumption that its y - and z -components of velocity were zero and that it suffered no collisions on the way. Actually, collisions do occur, and the molecules have y - and z -components of velocity. These may, however, be disregarded, since there are as many molecules of a certain velocity entering the cylinder from neighboring regions as are leaving the cylinder at any given instant.

Considering now the total number of molecules in the cylinder, we let the number per unit volume be N ; and since the volume of the cylinder is V_x , the total number of molecules is NV_x . Then, in Eq. (4-41), which gives the distribution of the velocity components, we replace n by NV_x to get the following expression for the number of molecules that have x -components of velocity between V_x and $V_x + dV_x$:

$$dn_{V_x} = \frac{NV_x}{\sqrt{\pi} c_0} e^{-\frac{V_x^2}{c_0^2}} dV_x \quad (4-67)$$

This number is also the number of molecules in this velocity range that reach the base $ABCD$ in 1 sec. To find the total number of molecules that cross the base per second, all that is necessary is to integrate over all possible velocities. Thus η , the number crossing unit area per second, is

$$\eta = \int_0^{\infty} \frac{NV_x}{\sqrt{\pi} c_0} e^{-\frac{V_x^2}{c_0^2}} dV_x$$

Upon evaluating the integral, it is found that this number is

$$\eta = \frac{Nc_0}{2\sqrt{\pi}} \quad (4-68)$$

But $c_0 = \frac{\sqrt{\pi}}{2} \bar{c}$; so η may be expressed as

$$\eta = \frac{N\bar{c}}{4} \quad (4-69)$$

Thus this formula shows that the number of molecules which cross a unit area per second is equal to the number of molecules contained in a rectangular volume of the gas having a unit cross section and a length one-fourth the arithmetical average speed of the molecules.

4-15. Evaporation of Electrons from Metals, or Thermionic Emission.—The phenomenon of thermionic emission of electrons from hot metals is accounted for in a general way as follows: The free electrons within the metal are conceived to have velocities of thermal agitation that depend on the temperature of the metal. The electrons that move outward through the surface of the metal and start to leave the surface experience attractive forces that tend to return them to the interior. The attractive force is mainly the electrostatic image force due to the charge of the electron itself. This surface attractive force is sufficiently large so that at ordinary temperatures the electrons are unable to leave the metal. As the temperature of the metal is raised the velocities and energies of the free electrons are increased until a portion of the electrons that arrive at the metal surface from within are able to overcome the surface force and escape.

Richardson¹ applied the molecular theory of gases to the problem of thermionic emission of electrons as early as 1901. His

¹ *Proc. Cambridge Phil. Soc.*, **11**, 286, 1901.

classical derivation is included here as an introduction to the subject, even though the resulting formula does not agree in detail with the modern theory. Richardson assumed that the perfect-gas laws may be applied to the free electrons in the metal and furthermore that the electrons are in thermal equilibrium with the metal and have a Maxwellian distribution of velocities. The surface attractive force is directed perpendicularly to the surface and is large at points very close to the surface but falls off rapidly with the distance. The electron that leaves the surface must give up part of its kinetic energy to overcome the attractive forces. This amount of energy is characteristic of the metal; it is called the *work function* of the metal and is represented by the letter W when expressed in joules. It is clear that only those electrons arriving at the surface of the metal with a component of velocity normal to the surface greater than a definite value will be able to escape. This value of velocity is easily expressed in terms of the work function; for if we take the x -axis perpendicular to the surface of the metal, then those electrons escape for which

$$\frac{1}{2}mV_x^2 > W \quad (4-70)$$

Here, m is the mass of the electron and V_x is its normal component of velocity when it reaches the surface. We let N designate the concentration of free electrons in the metal. We next apply Eq. (4-67) to give the number of electrons that arrive at the surface with x -components of velocity in the range from V_x to $V_x + dV_x$ per second.

$$dn_{V_x} = \frac{NV_x}{\sqrt{\pi} c_0} e^{-\frac{V_x^2}{c_0^2}} dV_x \quad (4-67)$$

According to the inequality (4-70) only those electrons having values of V_x equal to or greater than $\sqrt{2W/m}$ will escape. If we let η denote the number that escape per unit area per second, then we can find η by integrating Eq. (4-67) from the critical velocity up to infinity.

$$\eta = \int_{\sqrt{\frac{2W}{m}}}^{\infty} \frac{NV_x}{\sqrt{\pi} c_0} e^{-\frac{V_x^2}{c_0^2}} dV_x$$

or

$$\eta = \frac{Nc_0}{2\sqrt{\pi}} e^{-\frac{2W}{c_0^2 m}} \quad (4-71)$$

The most probable speed c_0 is characteristic of the temperature of the metal and according to Eq. (4-50) has the value $\sqrt{2kT/m}$. Upon introducing the value of c_0 into Eq. (4-71), we find

$$\eta = N \left(\frac{k}{2\pi m} \right)^{\frac{1}{2}} T^{\frac{1}{2}} e^{-\frac{W}{kT}} \quad (4-72)$$

The current density J carried by the emitted electrons is obtained at once by multiplying the number emitted per unit area per second by the electronic charge. Thus,

$$J = \eta q_e = N q_e \left(\frac{k}{2\pi m} \right)^{\frac{1}{2}} T^{\frac{1}{2}} e^{-\frac{W}{kT}} \quad (4-73)$$

Equation (4-73) may be written in the following form to emphasize the variation of the current density with the temperature:

$$J = a T^{\frac{1}{2}} e^{-\frac{b}{T}} \quad (4-74)$$

Here a and b are considered to be constants for a given metal, since N and W in Eq. (4-73) are usually assumed to be independent of the temperature. Because of the predominant influence of the exponential factor on the variation of the current density with the temperature, Eq. (4-74) agrees fairly well with thermionic-emission experiments. However, more recent theories have shown that the $T^{\frac{1}{2}}$ factor should be modified. The modern theory will be discussed in connection with the detailed study of thermionic emission (Secs. 6-2 and 6-3).

Problems

1. If a quantity of neon gas weighing 0.02 kg. is confined in a volume of 0.0012 cu. m. at a temperature of 50°C., what will be the pressure of the gas? Express the result in atmospheres and in millimeters of mercury.
2. What mass of gas would be required to fill a neon sign tube to a pressure of 8 mm. Hg at a temperature of 25°C. if the tube has a cross section of 40 sq. mm. and a length of 2 m.?
3. Helium gas has a density of 0.1785 kg./cu. m. at normal atmospheric pressure and at a temperature of 0°C. What is the r.m.s. average velocity of the helium atoms under these conditions?
4. (a) Calculate the r.m.s. velocity at a temperature of 1000°K. of the particles of each of the following gases on the assumption that they obey the laws of perfect gases: H_2 , N_2 , Hg , and a gas consisting solely of electrons. (b) What is the average energy of each particle? Express in electron volts.
5. A liter flask contains 32 g. of O_2 . What is the molecular concentration in the flask?

6. (a) Calculate the molecular concentration of nitrogen gas under a pressure of 10^{-5} mm. Hg and at a temperature of 80°C . (b) Under the conditions of part *a*, what is the average volume containing one molecule? (c) Would the results in *a* and *b* be altered if a different gas were used?

7. Assume that 0.4 g. of He and 1.4 g. of N_2 are mixed in a discharge tube having a volume of 10 liters. If the temperature is 27°C ., what is the pressure exerted by the mixture?

8. Prove that c_0 is the most probable speed of the molecules when the distribution of speeds is given by Eq. (4-42).

9. Consider 10 million nitrogen molecules confined in a volume and maintained at a temperature of 227°C . (a) Find the approximate number of these molecules that have speeds in the range from 599 to 601 m./sec. (b) Find the approximate number of the molecules that have *x*-components of velocity between +599 and +601 m./sec.

10. What is the ratio of the number of molecules in a small range of speeds dV taken at a speed of $2c_0$ to the number of molecules in the same range taken at c_0 , on the assumption that a Maxwellian distribution exists?

11. What fraction of the molecules in a given quantity of gas will have positive *x*-components of velocity greater than $2c_0$ at a given instant? HINT: The solution can be made to depend on the probability integral, tabulated values of which may be found in "A Short Table of Integrals," B. O. Pierce.

12. (a) Calculate the mean free paths of the molecules of each of the gases H_2 , N_2 , and CO_2 , if the temperature is 100°C ., the pressure is 10^{-3} mm. Hg, and the mean free paths have the values given in Appendix III under standard conditions. (b) What would be the mean free path of electrons in each of the gases in part *a*?

13. (a) The mean free path of the molecules of a certain gas at a temperature of 25°C . is 2.63×10^{-8} m. If the radius of the molecule is 2.56×10^{-10} m., find the pressure of the gas in millimeters of mercury. (b) Calculate the number of collisions per meter of path made by the molecule in part *a*.

14. Under certain conditions the molecules in a container have a mean free path of 2 mm. (a) What proportion of the given number of molecules will be traveling, at any instant, on paths longer than 20 mm.? (b) What proportion of the molecules will be traveling on paths in the range from 10 to 12 mm.?

15. If the mean free path of electrons in oxygen under standard conditions is 51×10^{-8} m., to how high a vacuum would one have to go in order that 90 per cent of the electrons emitted from an electron gun would strike a screen 0.15 m. away without making collisions with the oxygen molecules?

16. A cube measuring 0.1 m. on an edge contains N_2 gas under standard conditions of temperature and pressure. What will be the ratio of the number of molecules striking 1 sq. cm. of one wall per second to the total number of molecules in the cube?

17. Starting with Eq. (4-41), show that the average energy associated with one component of velocity is $\frac{1}{2}kT$ per molecule.

18. Show that the average energy, associated with the velocity components normal to a surface, of the molecules passing through this surface is kT per molecule. Why is this result different from that in Prob. 17?

19. Calculate the ratio of the number of electrons that escape from a metal surface to the number that arrive at the surface from the interior if the temperature of the metal is 2500°K. and its work function is 7.2×10^{-19} joule and if it is assumed that the electrons in the metal act as a perfect gas.

General References

- 4-1. BLOCH, E.: "The Kinetic Theory of Gases," 2d ed., Methuen & Co., Ltd., London, 1930.
- 4-2. JEANS, J. H.: "Dynamical Theory of Gases," 4th ed., Cambridge University Press, London, 1925.
- 4-3. LOEB, L. B.: "Kinetic Theory of Gases," 2d ed., McGraw-Hill Company, Inc., New York, 1934.
- 4-4. SLEPIAN, J.: "Conduction of Electricity in Gases," pp. 4-36, Educational Dept., Westinghouse Electric & Manufacturing Co., 1933.

CHAPTER V

CONDUCTION PROCESSES

5-1. Conductivity in the Normal State.—From the discussion in Chap. I, it is evident that when a gas exhibits conductivity it must possess a number of free charge carriers. These charge carriers can be either electrons or ionized molecules of the gas. The production of these ions from neutral atoms or molecules, however, requires a definite amount of work or energy; and, since any dynamical system tends to exist in that state for which the potential energy is lowest, the free electrons and ionized molecules will soon disappear. The various possible processes by means of which the ionization appears or disappears will be mentioned in the following sections and discussed in more detail in later sections. For the present, it should be sufficient to state that the average life of any particular charge carrier is very short. In the absence, then, of some process that continually replenishes the supply of free ions, the gas quickly loses all its conductivity.

We should expect that in the normal state all gases, even the vapors of metals, would be nonconductors. This is not strictly true, for even under the most favorable conditions all gases conduct electricity to some extent. Although this conductivity has been attributed to leakage over insulators, dust particles in the gas, and moisture in the gas, careful experiments have definitely shown that there is an appreciable leak through the gas.

Evidently some processes for the production of ionization are active at all times, regardless of the conditions under which the measurements are made. Tests under every kind of condition have always revealed this conductivity to a greater or less degree. The generally accepted opinion now is that this conductivity is produced by two causes which are practically impossible to eliminate. One cause is cosmic rays falling on the earth from interstellar space. The other is the disintegration of minute quantities of radioactive substances either in the current-meas-

uring device itself or in other solids near by. The natural conductivity produced by these agents is very small indeed. For example, in some cases the observed conductivity can be completely explained as due to the radiations produced by the disintegration of as few as 10 or 20 radioactive atoms per 100 sq. cm. of surface area per hour. This slight amount of ionization is sufficient, however, to initiate the breakdown of a sphere gap when the voltage across the gap is high enough. Without this ionization, breakdown would not occur except at extremely high voltages. Since clean, dry air has a resistivity of approximately 10^{16} ohm-cm., which is about the same as the resistivity of sulphur, we can, for all practical purposes, consider gases in the normal state as insulators.

5-2. Ionizing and Deionizing Agents.—Since all gases normally are insulators, it is obvious that when a gas becomes a good conductor, as it very often does, certain processes must be taking place that alter the amount of ionization present. Let us designate as *ionizing agents* all those means by which ionization can be made to appear in a gas. Conversely, let us call all those means by which ionization is made to disappear *deionizing agents*. We see, then, that with a gas in its normal state, as described above, the slight conductivity present can be attributed to the action of two ionizing agents, namely, cosmic rays and natural radioactivity. Under other conditions, other ionizing agents may be active—several examples will be discussed below. In any particular discharge, we shall always have one or more ionizing agents acting, as well as one or more deionizing agents. If this were not true, the amount of ionization present in the gas would continually increase or decrease. If an equilibrium condition is to be established, in which the net amount of ionization present does not change with time, we shall have to have as much ionization produced by the ionizing agents as is lost by the action of the deionizing agents.

The more important ionizing and deionizing agents may be classified as follows:

IONIZING AGENTS:

a. Cosmic Rays.—As mentioned above, these rays are continually falling on the earth from interstellar space. The rays possess an enormous amount of energy (enough to penetrate as

much as 3,000 ft. of solid rock, in some cases); consequently, they very readily ionize the air or gas through which they pass or liberate secondary electrons from the solids through which they pass. These secondary electrons may, in turn, produce more ionization.

b. Natural Radioactivity.—All substances are contaminated to a very slight extent with some of the heavy naturally radioactive elements. The amount present in many cases is so minute that even the most refined chemical analysis will not reveal its presence; yet the fact that high-energy radiations are given off from the material is an indication of the presence of radioactive elements. It is the radiations given off in the disintegration of these elements that produce ionization in the gas. In most problems in conduction in gases the ionization produced by this means or by cosmic rays is so small in comparison with that produced by other agents that it can be completely ignored. There are circumstances, however, when no other ionizing agents are acting, the initiation of a spark mentioned in the last section, for example. Then these ionizing agents may be important.

c. X Rays.—X rays produce ionization in a gas in much the same way as cosmic rays or the radiations from radioactive substances can produce ionization in the gas. Whenever they penetrate a gas, they leave behind them an ionized path. The amount of ionization produced by any beam of X rays depends strongly on the intensity and the energy (or wave length) of the X rays. The process involves the conversion of the electromagnetic energy of the X-ray beam into the potential and kinetic energy of an ionized atom or molecule. In this sense the process can be called a gaseous photoelectric effect as distinct from a surface photoelectric effect (see *h* below).

d. Intense Electric Fields (Ionization by Collision).—This agent requires the presence of some auxiliary ion source; but once a source of ionization has been obtained, the amount of ionization present can be made to increase very rapidly by the application of a strong electric field. The mechanism here involves the acceleration of the charged particles by means of the electric field until they have sufficient energy to produce more ionization in colliding with neutral atoms or molecules. These new ions, also, can then be accelerated until they, too, produce ionization by collision.

e. *High Temperatures (Thermal Ionization).*—If a gas is heated to a sufficiently high temperature, the neutral atoms themselves may possess enough kinetic energy so that in colliding with one another some ionization may be produced. The temperatures necessary for this, however, are extremely high, of the order of 5000 or 6000°C.; consequently, this form of ionization is encountered only under special conditions, for example, in a heavy-current arc in a high-pressure gas.

f. *Diffusion.*—Since the ionization present in a gas is due to atoms or molecules themselves being ionized, the charged particles will move about in the gas (in the absence of a strong electric field) in much the same way as the neutral atoms or molecules. Because of their thermal motions the ions can and will move into regions of low ionization and will increase the conductivity in these regions. Although the process of diffusion does not increase the total amount of ionization present, it does serve to increase the ionization in local regions, and in this sense it is an ionizing agent. *Convection* currents in the gas can be classed as a special type of diffusion process because its action is so similar to diffusion in producing ionization in some regions. In this process, all the molecules of the gas possess a large component of velocity in the same direction, and the gas as a whole then seems to move.

g. *Flames, Chemical Effects, etc.*—It has long been known that the gas in the neighborhood of a flame exhibits a very marked electrical conductivity. The flame itself thus produces a conducting state in the gas. It is also known that in many chemical reactions the gases involved become ionized. Other experiments show that air that has been passed over moist phosphorus or that has been bubbled through water is also ionized. Any one of these processes then could be called an ionizing agent, for its presence produces ionization and conductivity in the gas.

Up to now, all the ionizing agents we have been considering are those which can produce ionization throughout the body of the gas. There are certain other ionizing agents that are active only at surfaces. They are important, nevertheless; for, of course, any production of free charges at the boundary of a gas will increase the net number of free charges in the gas and consequently its conductivity. In many ways the following ionizing agents are more important than any of the preceding—excepting

perhaps the electric field—for at least one of these ionizing agents is encountered in practically every problem of gaseous conduction.

h. Photoelectric Emission.—All types of electromagnetic radiation possess a certain definite amount of energy depending upon the frequency or wave length of the radiation. When these rays, then, strike a surface or even the molecules of a gas, it is possible for the electromagnetic energy to be transformed into kinetic or potential energy in the molecules. The ability of long-wave-length radiation to produce heating is well known, and the ability of very short-wave-length (high-frequency) radiation to produce ionization in the gas was mentioned above. Light in the visible region or close to the visible region also possesses this ability. When light shines on some materials, it gives up its energy to electrons in the surface and some of these may thus gain a sufficient amount of energy to escape. The escape of electrons from a surface naturally increases the conductivity of the gas space adjacent to the surface.

i. Thermionic Emission.—It has been found that when metals are heated to a sufficiently high temperature, they begin to emit electrons. For most metals, this temperature is quite high—of the order of 2500°C .—but with some materials it may be as low as 1000 or 1200°C . Apparently the effect of the heat is to increase the kinetic energy of free electrons in the metal until some of them finally possess enough energy to escape from the surface. Under some conditions and with special materials, it is possible to observe the emission of positive molecular ions as well as electrons when heat is applied to a surface. Usually, however, the emission consists simply of electrons.

j. Intense Electric Field (High-field Emission).—When an intense electric field is produced right at the surface of a metal, we often find that electron emission takes place. The fields necessary for this to occur are very high—of the order of hundreds of millions of volts per meter—and consequently emission of this sort occurs only under special circumstances. The high electric field at the surface of a metal has the effect of reducing the amount of work that an electron would have to do in order to escape.

k. Secondary Emission.—High-speed electrons and positive molecular ions have the ability to cause the emission of secondary electrons from the electrodes or surfaces with which they collide.

In some types of discharge, it is easily possible for the insulated walls to acquire a large positive charge due to the bombardment of high-speed electrons and the emission of large numbers of secondary electrons. Here, again, the process is one in which internal electrons gain sufficient energy from the incident particles to escape from the surface.

DEIONIZING AGENTS:

a. Electric Field in the Gas.—Just as an electric field can be an ionizing agent by moving ions into a region, so it can also be a deionizing agent and move ions out of any particular region. Any charged particle in an electric field experiences a force, and movements of this sort are bound to occur. Ordinarily, one might expect the electric field to move as many ions into a given region as it moves out; but when the field distribution is very nonuniform, the electric field will often cause ions to disappear from the regions where the electric intensity is highest, thus producing a more nearly uniform field distribution.

b. Electric Field at Electrodes.—At electrodes in a discharge tube, the electric field removes ions from the gas space by moving them into the electrodes. When electrons are moved into the electrode, they pass directly into the metal; but when molecular ions strike the electrode, they are neutralized and return to the gas as neutral atoms or molecules either by extracting an electron from the metal in the case of positive molecular ions or by giving up their excess electrons to the electrode in the case of negative molecular ions.

c. Diffusion.—This is another process that can serve the dual role of either an ionizing or a deionizing agent. As a deionizing agent, it is one of the most important, appearing in practically every discharge. Wherever there is a region of high ion density, the normal thermal motions of the ions will result in a gradual decrease in ion density as the ions move out and neutral molecules move in.

d. Recombination in the Gas.—Ions of opposite sign have a natural affinity for one another—their mutual electrostatic field drawing them together. It is not surprising, then, that when their thermal movements bring them near enough to each other they should tend to be drawn together and recombine. This process can occur throughout the ionized gas, and the rate at

which it occurs is proportional to the numbers of each kind of ion present. As we shall see later, however, recombination occurs only when the negative ions are of molecular size. Electrons and positive ions rarely recombine, a fact indicating that this process is an important deionizing agent only at relatively high pressures— $\frac{1}{2}$ atm. or more—for only at pressures of this order do we find an appreciable number of negative molecular ions present in the gas.

e. Recombination at Surfaces.—When surfaces such as bounding walls are in regions of high ionic density, they may accelerate deionization. The process essentially is one in which the wall acts as a reservoir holding ions of one sign on its surface until

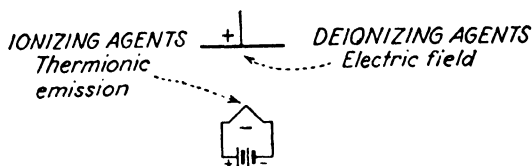


FIG. 5-1.—High-vacuum rectifier.

ions of the other sign have a chance to come up and combine with them to form neutral atoms or molecules. Unlike recombination in the gas, this process can go on at any pressure and is important at all pressures at which an appreciable number of positive ions is obtained. In equilibrium, we shall find equal numbers of positive and negative ions (or electrons) coming up to the wall and recombining there.

These, then, are the more important ionizing and deionizing agents. In order to see how they might enter the calculations for any particular discharge, let us examine a few typical discharges and see which agents would have to be taken into consideration.

Let us consider first an ordinary high-vacuum rectifier or triode such as is in every radio set. Here there is but one ionizing agent, namely, *thermionic emission*, and only one deionizing agent—the *electric field*. Incidentally, the electric field will appear as a deionizing agent in every discharge, for it is only by applying an electric field that we can measure a current. In this case, we might represent conditions as in Fig. 5-1.

If we now allow a little gas in the tube, the situation is altered considerably. This situation, which is encountered in tubes like

the 866 or in any thyratron, is represented in Fig. 5-2. Here we find two principal ionizing agents, namely, *thermionic emission* and the *electric field*, the latter producing ionization by collision. The deionizing agents are the electric field, as before, and, in addition, *diffusion* which causes ions to move out from the space between the electrodes and recombine on the walls. Hence,

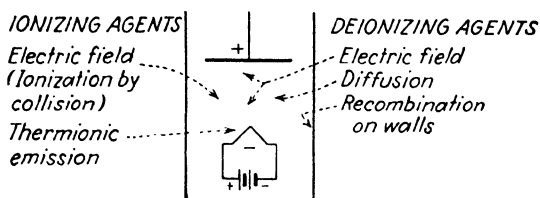


FIG. 5-2.—Gas-filled rectifier.

recombination on walls is also a deionizing agent. One might ask why diffusion and recombination on walls are not important in the case of the high-vacuum rectifier, also. It is true that initially some diffusion takes place to the walls; but since the number of positive ions present in this tube is insignificant, no recombination occurs at the walls, and the walls soon charge up to a high enough negative potential effectively to stop all further diffusion toward the walls.

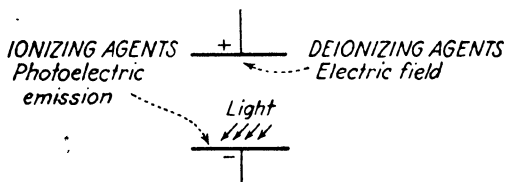


FIG. 5-3.—High-vacuum phototube.

Figure 5-3 represents the case of a high-vacuum phototube. The situation here is essentially the same as in the high-vacuum rectifier except that instead of thermionic emission as the ionizing agent we have *photoelectric emission*. As before, the electric field is the only deionizing agent.

The gas-filled phototube (Fig. 5-4) is analogous to the gas-filled rectifier; for here, also, we have ionization by collision due to the electric field in addition to the primary photoelectric emission. The deionizing agents are exactly the same, namely, diffusion in the gas, recombination on the walls, and the electric field at the electrode moving ions out of the gas space.

The case of a neon sign is only slightly different. In Fig. 5-5 the various ionizing and deionizing agents are indicated. We notice that in this case, as before, we have ionization by collision occurring in the gas; but now we do not have a cathode that is maintained in an emitting state by some external agency such as the thermionic or photoelectric cathode. Instead, the positive

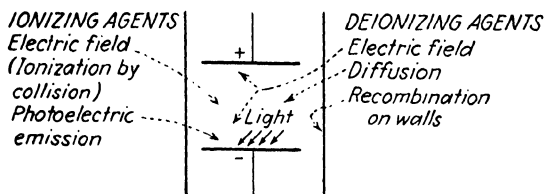


FIG. 5-4.—Gas-filled phototube.

ions generated by ionization by collision in the gas strike the cathode and liberate secondary electrons. In addition to the *secondary emission* that is occurring at the cathode, there is probably some *photoelectric emission* the light for which comes from the discharge itself. The electric field, diffusion, and recombination on the walls account for the deionization in this case, as in several of the preceding illustrations.

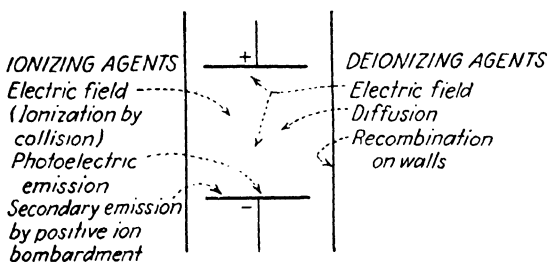


FIG. 5-5.—Neon sign.

The discharges discussed thus far all have had one principal characteristic in common—they are all low-pressure discharges. If we consider a high-pressure discharge such as an arc in air, we might represent the various ionizing and deionizing agents as shown in Fig. 5-6. In this case, we have electron emission from the cathode due to either a thermionic-emission effect or a high-field effect, depending upon the type of cathode material used—thermionic emission in the case of cathodes such as carbon or tungsten and high-field emission in the case of cathodes such as copper or aluminum. In either case, the energy for cathode

emission comes from the discharge itself. In the gas space, we have ionization due to the electric field as in other discharges but, in addition, a certain amount of *thermal ionization*; for, in the case of a high-pressure arc, the current densities and the energy input per unit volume of the arc are so high that the gas temperature easily reaches a temperature sufficient for thermal ionization. In regard to the deionizing agents, we see that again the electric field is active at both electrodes and diffusion is moving ions out of the discharge region; but now, since the walls are not necessarily close to the arc, we may have an appreciable amount of *recombination in the gas* as well as recombination on the walls.

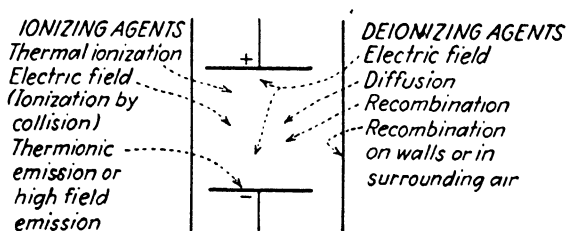


FIG. 5-6.—Arc in air.

These few illustrations should be sufficient to show the variety of combinations of ionizing and deionizing agents that are possible in any particular discharge. We shall leave it to the reader to analyze all the other illustrations that have, no doubt, occurred to him. There are two important facts that we must bear in mind in studying gaseous discharges. One is that in every discharge we shall have at least one ionizing agent and at least one deionizing agent acting. The other is that the characteristics of all discharges are determined by the kinds of ionizing and deionizing agents acting. Whenever we change the number or efficiency of the various ionizing or deionizing agents acting, we materially change the characteristics of the discharge. This latter point is especially important, for the design of a gas-discharge apparatus for any specific application is based upon it.

CHAPTER VI

ELECTRON EMISSION: THEORY AND TYPICAL APPLICATIONS¹

6-1. Thermionic Emission: Qualitative Observations and Theories.—The emission of electrons from hot bodies is a subject of considerable importance, not only because of the presence of the effect in certain arc discharges, but also because of the immense use of the phenomenon in radio and control tubes. The effect was observed by Edison as early as 1883 but was unexplained until the results of the brilliant investigations of O. W. Richardson and his coworkers were published starting in 1901. The view held by Richardson at that time, namely, that the electrons move out of the hot metal into the surrounding space in the same way that molecules of a liquid move out into the surrounding vapor, is still essentially correct. However, his view that the electrons inside the metal behave like a perfect gas in equilibrium with the metal atoms has given way to the treatment of the free electron atmosphere in the metal by means of the quantum-mechanical statistics developed by Fermi, Dirac, and others.

The factors involved in electron emission from hot cathodes could be studied experimentally by means of the *diode* tube and circuit shown in Fig. 6-1. Tests would show that the electron current I collected by the cold plate would be a function of the voltage E of the anode battery and of the temperature T of the cathode. In a general form, therefore,

$$I = f(E, T)$$

If cathodes of different materials, but of identical shape, were

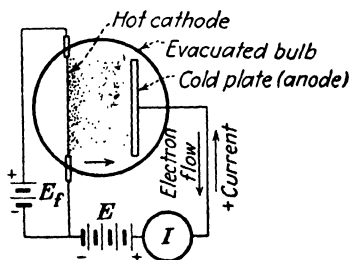


FIG. 6-1.—High-vacuum diode.

¹ General references for Chap. VI will be found on p. 224.

used, it would be found that the material of the cathode would have a profound influence on the current I in certain ranges of applied voltage E but would have little effect in other ranges.

Suppose that the temperature of the cathode in the tube in Fig. 6-1 is maintained constant at a value T_1 while the voltage E is varied. The curve $ABCD$ in Fig. 6-2 shows the relation

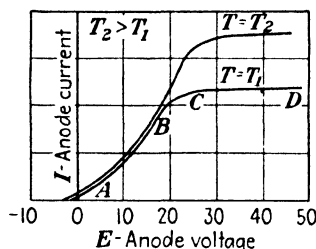


FIG. 6-2.—Voltage-current curves for a diode.

between E and I that results. One of the interesting features of this curve is the intercept with the vertical axis, which shows that some electrons reach the plate even though there is no accelerating electric field directed toward the plate. Another interesting feature of the curve is the *saturation region* from C to D . If a test is carried out at a higher cathode temperature T_2 , it will be found that the saturation current is higher and that a higher anode voltage is required to produce saturation as shown by the upper curve in Fig. 6-2. The natural inference is that in the saturation region, as CD , the anode voltage is high enough to cause all the electrons emitted from the cathode to be collected at the plate, whereas in the region AB a portion of the emitted electrons return to the cathode.

A qualitative explanation of the behavior exhibited in the range A to B can readily be developed by assuming the presence at any instant of a cloud of electrons between the cathode and plate as shown in Fig. 6-1. The electron cloud, or space charge, is considered to be sufficiently dense to cause the electric intensity just outside the cathode to urge the electrons back toward the cathode.

The postulated conditions can best be explained by reference to the curves in Fig. 6-3. The dotted curves show the variation of potential and intensity with no space charge in the gap, and the solid curves show the variations when sufficient

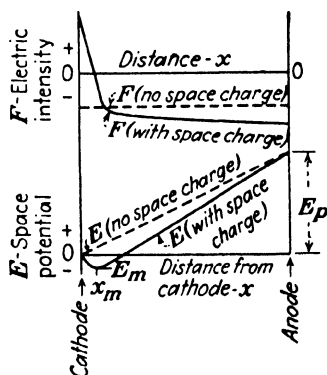


FIG. 6-3.—Curves showing the effect of space charge on the potentials and intensities in a parallel-plane diode.

space charge is present to reverse the intensity at the cathode. Under this condition the emitted electrons are in a retarding field until they reach the plane at x_m . At this point the electrons are in a region of minimum potential of the value E_m volts negative with respect to the cathode. Therefore, an electron leaving the cathode and already free of the surface attractive forces of the cathode would still need a kinetic energy at least equal to $E_m q_e$ in order to reach the point x_m and subsequently the plate. The general picture of the conditions near the cathode, therefore, should include the outgoing stream of electrons having velocities (and energies) distributed through a wide range from zero upward, and the incoming stream of electrons composed of those which failed to surmount the negative potential hump E_m . If a larger plate voltage E were impressed, it could be argued that x_m and E_m would decrease and that consequently the plate current would increase. If the plate voltage were increased to the point at which E_m becomes zero, then there would no longer be a retarding field at the cathode, and all the electrons that are emitted would reach the plate. This corresponds to the saturated condition as at C in Fig. 6-2. The complete mathematical analysis of the problem of the electron motions and concentrations as a function of the plate potential is difficult, but solutions based on simplifying assumptions will be discussed in Sec. 6-7.

Suppose now that the apparatus of Fig. 6-1 is used to investigate the variation of the plate current I as a function of the cathode temperature for a constant plate voltage E_1 . The resulting relation is shown graphically in Fig. 6-4. A second curve is shown for a higher anode voltage E_2 . In the ranges KH and KL in the curves an increase of anode voltage produces an insignificant increase in the current density. This is the region of saturation previously discussed. In the regions where the curves bend over toward the horizontal, it is evident that the voltage is not high enough to produce saturation; i.e., the space charge is the factor limiting the current.

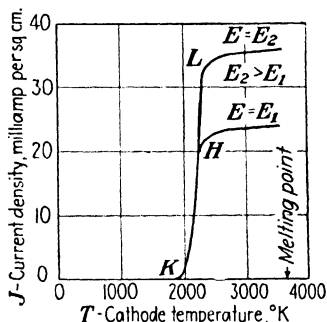


FIG. 6-4.—Current density at the cathode as a function of the cathode temperature for tungsten.

It is convenient to refer to the saturation portions, such as *C-D* in Fig. 6-2 and *KH* in Fig. 6-4, as conditions of *temperature-limited emission*; the current in situations in which the space charge is the predominant factor is called *space-charge-limited current*.

6-2. Theory of Electrons in Metals and of the Thermionic Emission of Electrons.—In Sec. 4-15 the free electrons in a metal were treated like the particles of a perfect gas, and the relation for the temperature-limited emission current was derived on the basis of Maxwell's distribution laws. Tests have shown that the emitted electrons have a Maxwellian distribution, a fact tending to confirm the assumptions concerning the electrons in the metal. However, an attempt to extend the electron theory of metals to explain their thermal properties, for example, their specific heats, led to failure, for the values of the specific heat showed that the free electrons did not share the heat energy in the same way as would a perfect gas in equilibrium with the metal atoms. Thus, although the classical theory could adequately describe the motions of the electrons in high vacuum, it was inadequate for the description of the motions of the electrons in a metal.

The classical theory is based on the assumption that the concentration of electrons is so small that the potential energy of two adjacent electrons by virtue of their electric charge at the average spacing is small compared with their average kinetic energy. This means that, if a potential energy of one one-hundredth of the kinetic energy is permitted, then at room temperature the concentration of electrons would be limited to 10^{16} and at 3000°K . the concentration would be limited to 10^{19} electrons/cu. m.¹ The concentration of electrons in the emitted current in high-vacuum tubes is within these limits,² and so the classical laws apply. If, however, we assume that one valence electron of each metal atom is sufficiently "free" to be a part of the electron gas, then we find that the electron concentration in the metal becomes so high³ that the assumptions underlying the

¹ KLEMPERER, OTTO, "Einführung in die Elektronik," p. 68, Verlag Julius Springer, Berlin, 1933.

² REIMANN, A. L., "Thermionic Emission," pp. 38-41, John Wiley & Sons, Inc., New York, 1934.

³ Strictly speaking, both the concentration and the mass of the particles

classical laws no longer are valid. For example, a unit volume (1 cu. m.) of tungsten, which has an atomic weight of 0.184 kg. and a density of 1.91×10^4 kg./cu. m., would contain 6.3×10^{28} electrons on the assumption of one free electron per atom.

The way out of the difficulty was found by Fermi, and at the same time by Dirac, who applied the Pauli exclusion principle to the statistics of groups of particles. The Pauli exclusion principle had previously been applied to the specification (or statement) of the permissible quantum states of an electron in an atom. Fermi and Dirac extended the application of the principle to the permissible energy states of the particles of a gas. The resulting statistical laws for electrons, together with their applications to the problem of thermionic emission, will be outlined in the following paragraphs,¹ and a comparison will be made between the Fermi-Dirac distribution laws and Maxwell's distribution laws.

In terms of the distribution of the speeds of the electrons, the Fermi-Dirac distribution law may be expressed as follows:

$$dn_V = \frac{8\pi m^3 n}{h^3 N} \left[\frac{V^2}{\exp\left(\frac{\frac{mV^2}{2} - W_i}{kT}\right) + 1} \right] dV \quad (6-1)$$

In this expression, dn_V is the number of electrons, out of a total number n , that have speeds between V and $V + dV$. N is the concentration of the electrons, h is Planck's constant equal to 6.62×10^{-34} joule-sec., and W_i is a constant. The value of the constant W_i must be consistent with the requirement that the integration of Eq. (6-1) over all values of V should give the total number of electrons, *i.e.*, n . Fermi showed that, to a first approximation,

$$W_i = \frac{h^2}{8m} \left(\frac{3N}{\pi} \right)^{\frac{2}{3}} \quad (6-2)$$

subject to the condition that $W_i \gg kT$. As an example to show that this condition is fulfilled for thermionic emitters, we calculate W_i for the electrons in tungsten on the assumption of one free electron per atom, *i.e.*,

$$N = 6.3 \times 10^{28}$$

determine the departure from the classical distribution laws. A gas that departs from the classical laws is called a *degenerate* gas. The electron gas in the metal is such a gas, but ordinary gases do not become degenerate even near their condensation point.

¹ For a detailed treatment, see K. K. Darrow, *Rev. Mod. Phys.*, **1**, 90-154, 1929. Also see general reference 6-1, Chap. V, and general reference 6-3.

Then, Eq. (6-2) gives $W_i = 9.23 \times 10^{-19}$ joule or 5.75 electron volts. Even at 3000°K . the value of kT is only 0.26 electron volt, and so the condition on W_i is satisfied.

One of the striking features of the Fermi-Dirac distribution law is that even at absolute zero the electrons are distributed through a range of energies. This is a complete departure from the classical theory and results from the occupancy by the electrons of all of the lowest permissible quantum-energy states. Inasmuch as the Pauli exclusion principle permits only a limited number of electrons to occupy each discrete energy level the available levels are filled to a rather high energy, resulting in the energy distribution at absolute zero. To examine the form of the distribution function at absolute zero, we let T approach zero in Eq. (6-1) and note that the exponen-

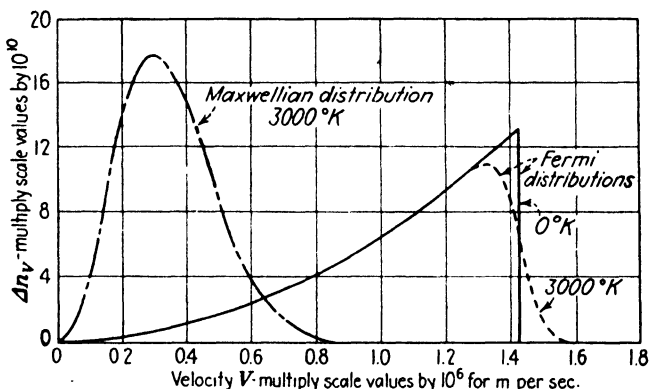


FIG. 6-5.—Fermi and Maxwellian distributions for the speeds of the electrons in 1 cu. mm. of tungsten, assuming $n = 6.3 \times 10^{19}$ electrons and $\Delta V = 10^{-3}$ m./sec.

tial term approaches zero or infinity depending on the value of $\frac{1}{2}mV^2$ relative to W_i . If $\frac{1}{2}mV^2$ is less than W_i , the exponential term is zero, and the distribution reduces to

$$dn_V = \frac{8\pi m^3 n}{h^3 N} V^2 dV \quad \begin{cases} T = 0 \\ \frac{1}{2}mV^2 < W_i \end{cases} \quad (6-3)$$

For values of $\frac{1}{2}mV^2$ greater than W_i the exponential term becomes infinite, and the right side of (6-1) is zero. Thus at absolute zero the distribution curve increases in proportion to the squares of the speeds according to (6-3) up to a certain maximum, then drops suddenly to zero. This behavior is shown by the solid curve in Fig. 6-5, which has been drawn for 1 cu. mm. of tungsten on the basis $N = 6.3 \times 10^{28}$; therefore $n = 6.3 \times 10^{19}$. The curve is calculated for finite increments ΔV and Δn_V by analogy with Fig. 4-8 (Sec. 4-11). The figure shows that the numbers of electrons in the range ΔV increases with V up to a maximum value at $V_m = \sqrt{2W_i/m}$, then drops suddenly to zero and remains at zero. We note that W_i is the energy corresponding to this maximum speed; hence, W_i is the maximum kinetic energy that an electron can have at absolute zero.

The effect of the temperature on the distribution of electron speeds can be shown by using Eq. (6-1) to calculate a curve for the velocity distribution in a particular case. The curve that results for tungsten at 3000°K. is shown by the dotted line in Fig. 6-5. It is seen that the effect of the temperature increase is to reduce the numbers of electrons having speeds slightly below V_m and to increase the speeds of some of the electrons to values above V_m , while most of the curve remains the same as at $T = 0^\circ\text{K}$.

Though the Fermi-Dirac laws apply to the electrons in metals, it is instructive to compare these laws with the corresponding Maxwellian laws. For this purpose the value of c_0 given by Eq. (4-50) is substituted into Eq. (4-42) to yield the following form of Maxwell's law:

$$dn_V = \frac{4n}{\sqrt{\pi}} \left(\frac{m}{2kT} \right)^{3/2} V^2 \exp \left(-\frac{mV^2}{2kT} \right) dV \quad (6-4)$$

This equation is used to plot the Maxwellian curve for electrons at 3000°K. The curve that results (Fig. 6-5) falls at much lower values of speed than the Fermi curve. Thus, Maxwell's law gives an average energy of the electrons far below that given by the Fermi law.

Although the curves in Fig. 6-5 show that the two distributions are actually quite different, an examination of the equations for the curves will show a marked similarity for large values of V . For example, if the value of $\frac{1}{2}mV^2$ is greater than about $1.1W_i$, then in Eq. (6-1) for the Fermi curve the exponential term is large compared with unity. Upon dropping unity from the denominator of the right-hand side of Eq. (6-1) the relation becomes

$$dn_V = C \left(\exp \frac{W_i}{kT} \right) V^2 \exp \left(-\frac{mV^2}{2kT} \right) dV \quad (6-5)$$

where C is a constant. It will be noticed that the variation with V in Eq. (6-5) is the same as for the Maxwellian distribution. However, the temperature variations are different for the two distribution laws. We shall show in what follows that the energy distribution of the electrons emitted thermionically as given by the Fermi laws is also in agreement with the Maxwellian energy distribution but that the temperature variation of the emission is different in the two cases.

The thermionic emission from a metal can be calculated by the same method as that used in Sec. 4-15 if the distribution of velocity components is known. Nordheim¹ has shown that according to the Fermi-Dirac laws the number of electrons in a unit volume having velocity components in the range dV_x is given by

$$dn_{V_x} = \frac{4\pi m^2 kT}{h^3} \log \left[1 + \exp \frac{-(\frac{1}{2}mV_x^2 - W_i)}{kT} \right] dV_x \quad (6-6)$$

The number of electrons dn_{V_x} that strike unit area of a surface per second with velocity components in the range dV_x will be the number dn_{V_x} given by

¹ *Physik. Z.*, **30**, 177, 1929.

Eq. (6-6) multiplied by V_z (see Sec. 4-14).

$$d\eta_{V_z} = \frac{4\pi m^2 kT}{h^3} V_z \log \left[1 + \exp \left(\frac{W_i - \frac{1}{2}mV_z^2}{kT} \right) \right] dV_z \quad (6-7)$$

The direction of the x -axis is taken to be normal to the surface. To simplify the discussion we can restate the distribution given by Eq. (6-7) in terms of the kinetic energy W_z associated with the x -component of velocity, *i.e.*, in terms of

$$W_z = \frac{1}{2}mV_z^2$$

Then Eq. (6-7) transforms into the following:

$$d\eta_{W_z} = \frac{4\pi mkT}{h^3} \log \left[1 + \exp \left(\frac{W_i - W_z}{kT} \right) \right] dW_z \quad (6-8)$$

To examine this distribution function, we first inquire as to its form at absolute zero. In the range where $W_z < W_i$, as T approaches zero, the

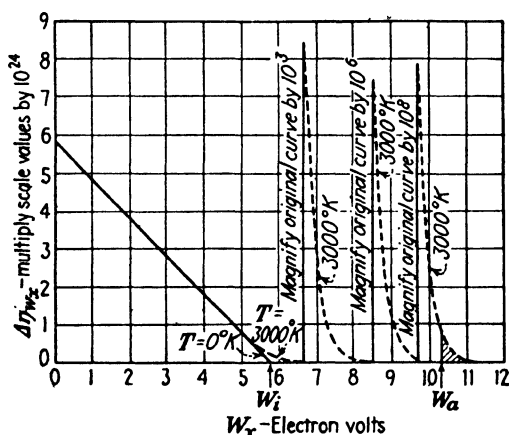


FIG. 6-6.—Fermi energy distributions for “normal” energies of electrons crossing a surface in tungsten. $\Delta W_z = 10^{-3}$ e.v.

exponential term in (6-8) becomes large compared with unity. Upon dropping unity from the bracketed term, the equation reduces to

$$d\eta_{W_z} = \frac{4\pi m}{h^3} (W_i - W_z) dW_z \quad \begin{cases} W_z < W_i \\ T = 0 \end{cases} \quad (6-9)$$

Thus $d\eta_{W_z}$ is a linear function of W_z of negative slope for $W_z < W_i$. For values of $W_z > W_i$, the logarithmic function in (6-8) approaches zero as T approaches zero, and the whole function approaches zero. Thus the curve for $T = 0$ is a straight line of negative slope which halts abruptly at the axis of abscissas at $W_z = W_i$ (see Fig. 6-6). This conforms with the interpretation of W_i as the maximum kinetic energy that an electron can have at absolute zero.

When Eq. (6-8) is used to plot the distribution in energy of the electrons that cross a surface in tungsten at an elevated temperature (3000°K.), we

find a remarkably small change in the curve. The curve no longer terminates at $W_z = W_i$ but instead extends out to higher values of W_z (Fig. 6-6). The successive magnifications of the curve shown in the figure permit an extension of the curve out to a W_z of about 11 e.v. It is these higher energy electrons that are involved in thermionic emission.

To show the distribution of the higher energy electrons more clearly the curves for $T = 0$ and $T = 3000^\circ\text{K}$. have been plotted to a logarithmic scale of ordinates in Fig. 6-7. For comparison, a curve for the corresponding

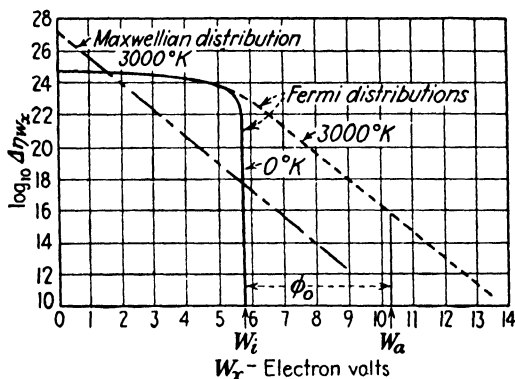


FIG. 6-7.—Fermi and Maxwellian energy distributions for electrons crossing a surface. $\Delta W_z = 10^{-3}$ e.v.

Maxwellian distribution is also shown. This is obtained from Eq. (4-67) by changing the variable to W_z , giving

$$d\eta_{W_z} = \frac{N}{(2\pi mkT)^{\frac{3}{2}}} \exp\left(-\frac{W_z}{kT}\right) dW_z \quad (6-10)$$

It is at once apparent from Fig. 6-7 that the higher energy electrons in the Fermi-Dirac distribution have the same distribution in energy as those which follow the Maxwellian distribution law. To confirm this observation, we can examine the form of Eq. (6-8) for values of W_z greater than about $1.2W_i$. For all practical cases the term $\exp\left(-\frac{(W_z - W_i)}{kT}\right)$ in Eq. (6-8) will be a small quantity, for kT is small compared with W_i . Then it is possible to apply the following series

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots \quad \text{where } x^2 < 1$$

to the log function in Eq. (6-8). Since the terms decrease very rapidly, we save only the first term, with the result that to a high degree of approximation

$$d\eta_{W_z} = \frac{4\pi mkT}{h^3} \exp\left(\frac{W_i}{kT}\right) \exp\left(-\frac{W_z}{kT}\right) dW_z \quad (6-11)$$

Upon comparison of (6-11) with (6-10), it is seen that when the temperature is constant both have the same form of variation with W_z but that the functions of the temperature are different in the two expressions. It is not difficult to show¹ that the Fermi theory predicts the emission of electrons whose energies have a Maxwellian distribution.

The method already used in Sec. 4-15 can be used to apply the Fermi theory to the calculation of the thermionic emission. Let the x -axis be normal to the surface, and let W_a be the value of $\frac{1}{2}mV_x^2$, i.e., the "perpendicular" energy that an electron needs before it can just overcome the surface attractive forces. Clearly, W_a must be larger than W_i ; otherwise, the metal would be losing electrons even when cold. If W_a is greater than $1.2W_i$, then the approximate equation (6-11) will give the distribution of the electrons involved in the emission problem. The number being emitted per unit area per second equals the integrated value of Eq. (6-11) from the critical energy W_a to infinity, thus:

$$\eta = \frac{4\pi mkT}{h^3} \exp \frac{W_i}{kT} \int_{W_a}^{\infty} \exp \left(-\frac{W_z}{kT} \right) dW_z \quad (6-12)$$

This integrates directly to give

$$\eta = \frac{4\pi mk^2 T^2}{h^3} \exp \left(-\frac{W_a - W_i}{kT} \right) \quad (6-13)$$

This number can be interpreted in terms of the Fermi-Dirac curve in Fig. 6-6 as being all the electrons represented by the shaded area under the curve from W_a out to infinity.

Multiplication of Eq. (6-13) by the charge on an electron, q_e , gives the saturation current density of the emitted electrons

$$J_s = \frac{4\pi mk^2 q_e}{h^3} T^2 \exp \left(-\frac{W_a - W_i}{kT} \right) \quad (6-14)$$

This is usually written in the simpler form

$$J_s = A_0 T^2 \exp \left(-\frac{\phi_0 q_e}{kT} \right) \quad (6-15)$$

¹ At a constant temperature, Eq. (6-11) may be written

$$d\eta_{W_z} = C \exp \left(-\frac{W_z}{kT} \right) dW_z$$

W_z must exceed W_a before any electrons are emitted. However, the emitted electrons have an energy $(W_z - W_a)$, since they lose an energy W_a at the boundary. The emitted electrons therefore have a distribution in energy given by

$$d\eta_{W_z} = C \exp \left(\frac{W_a - W_z}{kT} \right) dW_z \quad \{W_a < W_z\}$$

This is of the same form of energy distribution as the Maxwellian distribution [Eq. (6-10)], but the number of emitted electrons is different from that given by the classical laws.

where

$$A_0 = \frac{4\pi mk^2 q_0}{h^3} \quad (6-16)$$

and

$$\phi_0 q_0 = W_a - W_i \quad (6-17)$$

When Eq. (6-15) is compared with the result based on Maxwell's laws [Eq. (4-74)], the difference in the temperature factor is obvious. The predominant influence of the temperature, however, is contained in the exponential term, which is the same in both equations. The expression for the thermionic-emission current (6-15) has also been derived on the basis of thermodynamical arguments by Richardson and Dushman, and the latter obtained a value of A_0 in 1923 which was the same as that given by Eq. (6-16) except for a factor of $\frac{1}{2}$. The factor of $\frac{1}{2}$ arose because Dushman did not take into account the electron spin, which was, however, unknown at that time.

The value of the universal constant A_0 as given by Eq. (6-16) is 1.204×10^6 amp./(m.^2)(deg.²), or 120.4 amp./(cm.^2)(deg.²). The theory outlined above is unable to predict values of the quantity ($W_a - W_i$) for a particular metal. The value of W_a , we recall, is the critical normal energy needed for an electron to escape from the surface forces, and W_i is the maximum energy of any of the electrons at $T = 0^\circ\text{K}$. The difference ($W_a - W_i$) is called the *net work function* of the metal. Thus, we see by studying Eq. (6-15) that ϕ_0 is the *net work function* expressed in electron volts. The quantity W_a is designated the *outer work function* and W_i the *inner work function*. By comparison with the early Richardson equations in Sec. 4-15, we note that the work function W defined there is analogous to W_a , the outer work function, but that it occupies the same position in the earlier emission equation as the net work function ($W_a - W_i$) does in Eq. (6-14).

TABLE 6-1

Metal	Number of electrons per atom (assumed)	Inner work function W_i (calculated), e.v.	Net work function (observed), e.v.	Outer work function W_a (calculated), e.v.
W	1	5.75	4.5	10.25
W	2	9.1	4.5	13.6
Ni	2	11.7	5	16.7
Ni	3	15.3	5	20.3
K	1	2.1	1.8	3.9

The net work function can be measured by experimental determinations of the thermionic current (see Sec. 6-3). The value of the inner work function W_i depends on the number of free electrons per unit volume according to Eq. (6-2). Although this number of electrons is not known with cer-

tainty, there is evidence that indicates that in most metals there are one to three free electrons per atom. A few calculated values of W_i and observed values of ϕ_0 , together with the value of W_a that would be required to satisfy Eq. (6-15), are given in Table 6-1. To facilitate comparisons, the work functions are all expressed in electron volts. This table shows the order of magnitude of the values obtained. Experiments on the diffraction of beams of electrons impinging on the surfaces of metallic crystals, as originated by Davisson and Germer¹ and subsequently carried out by Rupp² and others, lead to independently determined values of W_a . For nickel, for example, Davisson and Germer find $W_a = 18$ volts and Rupp finds $W_a = 16.5$ volts. These experiments do not give great accuracy in the values of W_a , but the results obtained are in satisfactory agreement with the values in Table 6-1 for the assumption of two free electrons per atom.

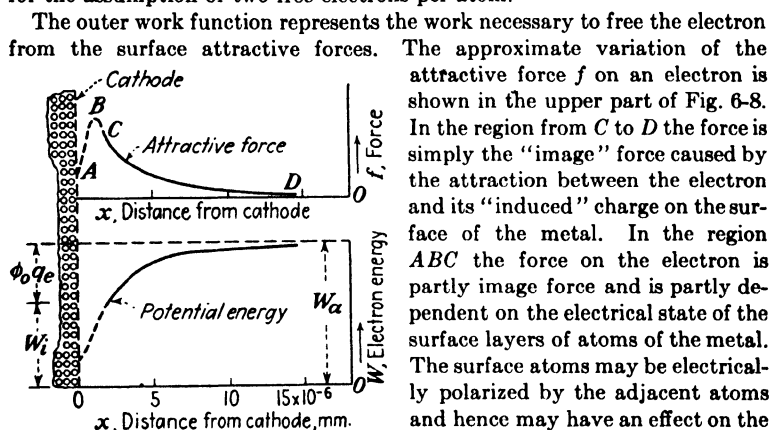


FIG. 6-8.—Attractive force and potential energy curves for an electron near a metallic cathode.

The lower curve in Fig. 6-8 shows how the potential energy of an electron would vary if calculated from the work done against the attractive forces in carrying an electron from the inside of the metal to the point in question, i.e., by $W = \int_0^x f dx$. The asymptote W_a represents the outer work function. Since an important portion of the force curve (ABC) depends on the kind of metal, the value of W_a would be expected to vary from metal to metal.

In Fig. 6-8 the inner work function has been plotted as the ordinate W_i . Thus the net work function $\phi_0 q_e$ can be shown in the diagram as the difference between W_a and W_i . The electrons which are thermally emitted are those which approach the surface with an energy greater than W_a .

¹ *Phys. Rev.*, **30**, 705, 1927.

² *Ann. Physik.*, **1**, 801, 1929; **5**, 453, 1930.

6-3. Temperature-limited Emission.—To make an experimental test of Eq. (6-15), it is necessary to measure the temperature-limited current emitted from a thermionic cathode as a function of temperature and then to test the data for conformity with the equation. The equation is altered slightly for this purpose by replacing $\phi_0 q_e/k$ by a new constant b_0 , thus:

$$J_s = AT^2 \exp \left(-\frac{b_0}{T} \right) \quad (6-18)$$

Then b_0 is related to ϕ_0 by the following:

$$b_0 = \frac{q_e}{k} \phi_0 = 11,600 \phi_0 \quad (6-19)$$

where ϕ_0 is the net work function in electron volts as defined in the preceding section.

The values of the constants A and b_0 in Eq. (6-18) have been determined experimentally for various clean metals under high-vacuum conditions. These constants can easily be evaluated from observations of cathode temperature and saturation current density by a graphical method. Thus Eq. (6-18) can be transformed into

$$\log_{10} J_s - 2 \log_{10} T = \log_{10} A - \frac{b_0}{2.303T} \quad (6-20)$$

Therefore, if $(\log_{10} J_s - 2 \log_{10} T)$ is plotted as a function of $1/T$, a straight line should result that has an intercept of $\log_{10} A$ and a slope of $-b_0/2.303$. This straight line is often called a Richardson line.

The experiments are very difficult to make in such a manner that accurate and consistent values of A and b_0 are obtained. The greatest difficulty lies in eliminating adsorbed gases and other impurities from the metal surfaces. In many cases the temperature range over which the emission current can be measured is rather limited. As a result of these difficulties the values of the thermionic constants reported by different workers cover a range of values. The following table of values for clean metals has been taken from the book by Reimann¹ except as noted:

¹ General reference 6-1.

TABLE 6-2.—THERMIONIC CONSTANTS OF CLEAN METALS

Metal	A , amp./($m.^2$) (deg. ²)	ϕ_0 e.v.	Melting point, deg. K.	Calculated emission at melting point (amp./sq. m.)
Platinum.....	$\left\{ \begin{array}{l} 17,000 \times 10^4 \\ 170 \times 10^4 \\ 32 \times 10^4 \end{array} \right.$	6.27^* $5.4 \uparrow$ $5.32 \dagger$	2028	4.6×10^{-3}
Nickel.....	$1,380 \times 10^4$	5.03	1725	8.4×10^{-3}
Tungsten.....	60×10^4	4.53	3655	456×10^4
Carbon.....	30×10^4	4.34§		
Copper.....	65×10^4	4.33	1356	9.8×10^{-5}
Molybdenum.....	55×10^4	4.15	2895	27.7×10^4
Tantalum.....	60×10^4	4.1	3300	361×10^4
Thorium.....	70×10^4	3.38	2118	2.88×10^4
Barium.....	60×10^4	2.11	1123	2.72×10^3
Cesium.....	162×10^4	1.81	299.4	5.1×10^{-20}

* DU BRIDGE, L. A., *Phys. Rev.*, **32**, 961, 1928.

† VAN VELZER, H. L., *Phys. Rev.*, **44**, 831, 1933.

‡ WHITNEY, L. V., *Phys. Rev.*, **50**, 1154, 1936.

§ REIMANN, A. L., *Proc. Phys. Soc. (London)*, **50**, 496, 1938.

Additional data on the work function of various metals have been compiled by Becker.¹

We note that the values of A in Table 6-2 vary considerably from the theoretical value of 120.4×10^4 derived in the preceding section. One explanation of the differences has been offered on the basis of the reflection of the electrons at the interface between metal and vacuum, in accordance with the wave theory of the electron. The result would be a reduction in the electron current and a reduction in the value of the constant A . The experimental values of A are, however, both smaller and larger than A_0 ; and this fact, together with theoretical calculations of the probable electron reflection,² lead to the conclusion that this effect is insufficient to explain the difference observed. Another explanation is found in the possibility of a variation of the work function with the temperature. An increase of the work function with the temperature would lead to a value of A smaller than A_0 , whereas a decrease would produce the opposite effect. The work function would need to increase only about 6 parts per

¹ General reference 6-3.

² General reference 6-1, pp. 30, 252.

100,000 per degree to reduce A from 120×10^4 to 60×10^4 . There is some evidence¹ that the temperature variation of the work function is of the right order of magnitude to account for the observed values of A .

The table above also gives the melting point and calculated electron emission at the melting point for the various metals. These figures show that tungsten, molybdenum, and tantalum give high emissions at their melting points. Of these, tungsten is the metal that has been used for the cathode in practical radio tubes. At present the use of tungsten cathodes is largely limited to high-power radio tubes, high-voltage high-vacuum rectifier tubes, and hot-cathode X-ray tubes.

6-4. Electron Emission from Composite Surfaces: Practical Emitters. *a. Thorium on Tungsten and Other Atomic-film Emitters.*—An important class of emitters is that in which a metallic core is covered with an extremely thin film of an electro-positive metal. This type of surface was discovered by Langmuir and Rogers² in 1913. They found that a tungsten filament containing from 1 to 2 per cent of thoria (ThO_2) gave a high electron emission after suitable heat-treatment. Langmuir later showed that the wire surface was partly or wholly covered with a monatomic layer of adsorbed thorium atoms or ions. The electron emission was found to depend considerably on the state of the surface, especially on the amount of thorium on the surface. The emission increased as the amount of thorium on the surface increased; but a maximum emission was soon reached, and further addition of thorium to the surface gave a decreased emission. Similar results were later obtained with thorium on molybdenum, and with the alkali metals deposited by evaporation onto tungsten. Values of the emission constants in Richardson's equation [Eq. (6-15)] for various surfaces corresponding to the condition of maximum emission are given in the following table. The values obtained in different experiments are subject to considerable variation; for example, Nottingham³ obtained values of ϕ_0 ranging from 2.38 to 2.83 for thoriated tungsten, with corresponding values of A ranging from 0.56 to 10.2 amp./(cm.^2)(deg.²).

¹ REIMANN, A. L., *Proc. Roy. Soc.*, **163A**, 499, 1937. BECKER, J. A., and W. H. BRATTAIN, *Phys. Rev.*, **45**, 694, 1934.

² *Phys. Rev.*, **4**, 544, 1914.

³ *Phys. Rev.*, **41**, 793, 1932.

TABLE 6-3*

Surface	A , amp./ (cm. ²)(deg. ²)	ϕ_0 , e.v.
Thorium on tungsten (W-Th).....	3.0	2.63
Thorium on molybdenum (Mo-Th).....	1.5	2.59
Barium on tungsten (W-Ba).....	1.5	1.56
Cesium on tungsten (W-Cs).....	3.2	1.36
Barium on oxygenated tungsten (W-O-Ba).....	0.18	1.34
Cesium on oxygenated tungsten (W-O-Cs).....	0.003	0.72

* General reference 6-1, Chap. III.

The procedure in the production of a thoriated tungsten cathode is essentially as follows: The tube is first prepared in the usual way. The glass bulb is degassed by baking while the vacuum pumps are connected, and the electrodes and filament are also outgassed by heating. A layer of magnesium or other "getter" is then deposited on the glass wall by evaporating a small pellet within the tube. The tube is sealed off immediately after thus removing traces of oxygen and other active gases. Then the filament,¹ which contains about 1.5 per cent of ThO₂, is heated to 2800°K. for 1 or 2 min. This causes the reduction of a small percentage of the ThO₂ to metallic Th. If any metallic thorium appears on the surface during this process, it quickly evaporates and is lost. The second operation consists in holding the filament at a temperature of 2100 to 2200°K. for a period of about 15 to 30 min. This is known as the activation period, for tests show that the electron emission increases during this period and approaches a steady value. The filament is now fully activated and is ready for use. At an operating temperature of about 1900°K. the emitting surface is quite stable, and the filament may be operated for long periods of time.

The investigations² of Langmuir and Becker and their coworkers have thrown considerable light on the physical

¹ The tungsten wire is usually given a preliminary heat-treatment in a hydrocarbon vapor such as naphthalene. This results in a chemical change of some of the outer surface to tungsten carbide. Such a "carbonized" filament has improved characteristics in that the evaporation of thorium is less than for the ordinary filament (see general reference 6-6, p. 34).

² General reference 6-1, Chap. III; general reference 6-5, Chap. IV. BRATTAIN, W. H., and J. A. BECKER, *Phys. Rev.*, **43**, 428, 1933. LANGMUIR, I., *J. Franklin Inst.*, **217**, 543, 1934.

processes involved in the activation and on the surface condition of this emitter. The processes involved include (1) reduction of ThO_2 , (2) "diffusion" of metallic thorium to the surface, (3) migration of thorium over the surface, and (4) the removal of thorium from the surface by evaporation. The reduction of thorium takes place at temperatures over 2600°K . The manner in which the thorium comes to the surface during activation has been made clear by recent studies of the surfaces by the electron microscope.¹ In this method the electrons emitted from a small area of the surface to be investigated are accelerated through an electron-lens arrangement (see Sec. 3-8) and focused on a fluorescent screen or a photographic plate. In this way a detailed observation can be made of the sources of the electrons. Bruche and Mahl found that the activation was accompanied by the appearance of thorium at discrete points on the surface of the tungsten. Ahearn and Becker found that an activation at 1500 to 1800°K . also resulted in the arrival of the thorium at discrete points on the surface; the boiling out of the thorium was termed a thorium "eruption." The electron images showed that the thorium spread out on the surface and the surface density of thorium decreased as the film spread out. By comparing the electron images with photomicrographs, they developed the hypothesis that the thorium is distributed through the tungsten crystals in small pores. These pores are conceived to deliver approximately all their thorium to the surface when an opening to the surface develops. The eruptions were observed on the tungsten crystal faces but appeared about equally often on the grain boundaries. This observation agrees with previous experiments which showed that a fine-grained structure of the filament facilitated activation.

Although there is still some disagreement, it appears that the fully activated surface corresponds to a monatomic layer of thorium atoms or ions which covers about 70 per cent of the surface. The thorium is held on the surface with forces so strong that appreciable evaporation does not begin until a temperature of about 2100°K . is reached. However, the thorium has some freedom to migrate over the surface down to as low a temperature as 1600°K . This was proved by an experiment of

¹ BRUCHE, E., and H. MAHL, *Z. tech. Physik*, **16**, 623, 1935. AHEARN, A. J., and J. A. BECKER, *Phys. Rev.*, **54**, 448, 1938.

Brattain and Becker¹ in which thorium was evaporated from an external source and allowed to deposit on one side of a flat tungsten ribbon. After 30 hr. operation at 1600°K. the electron emission from *both* sides of the ribbon was nearly equal, showing that the thorium had spread nearly uniformly over the surface.

In the range of temperature from 2000 to 2300°K. the rate of arrival of thorium on the surface is more rapid than the evaporation, and hence this range may be called the *activating* range. If a filament is held at an activating temperature, the surface layer of thorium builds up until an approximately monatomic

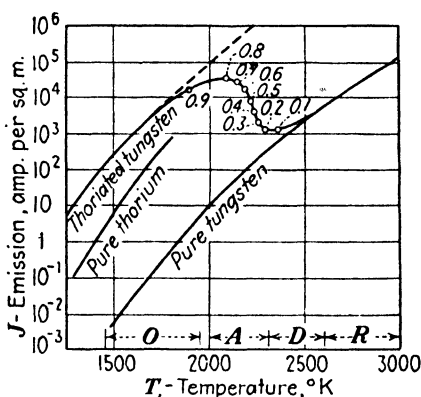


FIG. 6-9.—Electron emission from thoriated tungsten, thorium, and tungsten. O indicates the operating temperature range, A the activating, D the deactivating, and R the reducing ranges. (After Bruche, E., and H. Mahl, *Z. tech. Physik*, **16**, 624, 1935.)

layer is reached. The arrival of additional thorium on the surface would tend to produce deeper layers, but the outer atoms of deeper layers evaporate much more readily than the first layer. Thus an equilibrium surface condition is reached which corresponds to a nearly monatomic layer. Above 2300°K. the evaporation of thorium is usually more rapid than the replenishment, so that a deactivation ensues. Below 2000°K. the rates of arrival and evaporation are both small, but a slight preponderance of diffusion in the operating range helps to maintain the surface. The various ranges of temperature are related to the emission from tungsten and tungsten partly or wholly covered with thorium in Fig. 6-9. The numbers on the transition curve between W-Th and pure W refer to the portion of the surface

¹ *Loc. cit.*

of tungsten covered with thorium expressed as a fraction of the covering for maximum emission.

It is interesting to note that the work function of a thoriated filament (2.6 volts) is much below that of tungsten (4.5 volts) and is below that of thorium itself (3.4 volts). Thus the emission from thoriated tungsten is higher than that from thorium (Fig. 6-9). The marked reduction of the work function must be caused by the effect of the electropositive thorium layer on the electrons as they emerge from the surface. Thus the forces on

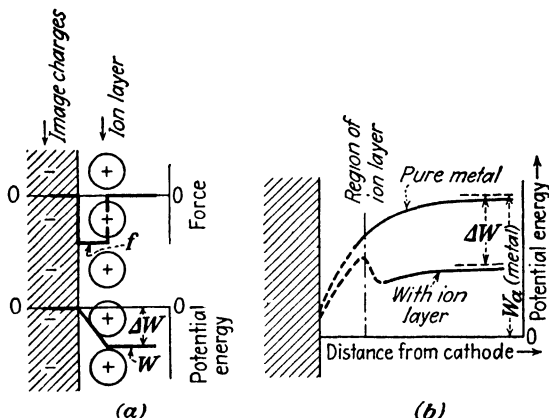


FIG. 6-10.—(a) Idealized force and energy curves caused by an ion layer. (b) Schematic diagram showing the effect of an ion layer on the potential energy of an electron near the surface.

the electrons caused by the thorium layer must be predominantly outward, *i.e.*, opposite to the image forces. Such forces could be accounted for if at least a part of the adsorbed thorium is ionized, a view held by Becker and also by de Boer.¹ An idealized picture of the conditions is obtained if the layer of ions is considered as a uniform, extended sheet of positive charge, with a corresponding sheet of negative charge on the metal surface. An electron passing through such a layer from left to right will experience a force f and will lose an amount of potential energy ΔW as illustrated in Fig. 6-10a. This will tend to reduce the work function of the surface as shown in Fig. 6-10b. The assumed curve for the base metal is here reproduced (see Fig. 6-8), and it is assumed that the ion layer is situated in a region of strong image forces.

¹ General reference 6-12, p. 110.

An alternative view regarding the condition of the surface was advanced by Langmuir and is held by Reimann.¹ According to this view the adsorbed atoms are neutral but electrically distorted, so that, on the average, the electrons are closer to the base metal than the positive nucleus. Thus a dipole is formed with its positive end outward. A layer of such dipoles will act to reduce the work function of the surface in much the same way as explained above for a layer of ions.

The behavior of thin layers of barium on tungsten is very similar to that of thorium on tungsten. Layers of cesium on

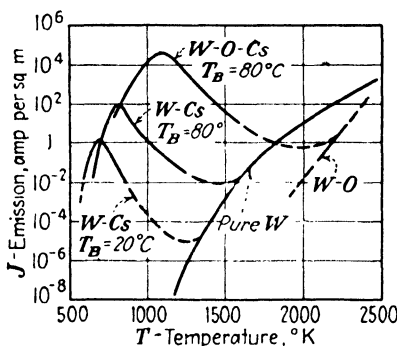


FIG. 6-11.—Electron emission from tungsten and oxygenated tungsten in the presence of saturated cesium vapor, and from pure tungsten and oxygenated tungsten. T_b is the bulb temperature. (Langmuir, I., *Ind. Eng. Chem.*, **22**, 392, 1930.)

tungsten behave in a somewhat similar way, also, but with an added complication. Since metallic cesium melts at $26^\circ C$., it has a high vapor pressure at room temperatures. Therefore, a bulb containing cesium metal will contain cesium vapor whose atomic concentration depends markedly on the bulb temperature. Thus the rate of arrival of the cesium atoms on the filament surface will also depend on the bulb temperature. The net result is that the thermionic properties of the filament depend not only on the filament temperature but on the bulb temperature as well. This is brought out in the curves in Fig. 6-11.

Though a monatomic layer of oxygen reduces the thermionic emission of tungsten by a large factor, a surface prepared by depositing cesium on the oxygenated tungsten has a very high emission. Furthermore, the affinity of the surface for the

¹ General reference 6-1, pp. 180ff.

cesium atoms is increased, so that the filament may be operated at a higher temperature. Surfaces similar to W-O-Cs are not important industrially as thermionic emitters but find their main use as photoelectric and secondary-emission cathodes (see Secs. 6-10 and 6-13).

b. Oxide-coated Electron Emitters.—The development of the oxide-coated cathode started with the discovery by Wehnelt¹ in 1903 that the emission from oxides of the alkaline earth metals, for example, barium oxide or calcium oxide, was copious even at relatively low temperatures. It was not until about 1914, however, that the Western Electric Company began to make tubes with oxide-coated cathodes for use in the telephone industry. At present, practically all receiving types of radio tubes have the oxide-coated cathodes.

The many variable factors entering into the preparation of oxide-coated cathodes results in corresponding differences in manufacturing techniques. In one method a ribbon or wire of nickel or nickel alloy is coated with a water suspension of a mixture of barium and strontium carbonates and is then subjected to a preliminary heating in an atmosphere of carbon dioxide. This coating process is repeated until the desired thickness is obtained. The filament is next mounted in the tube, and the evacuation is begun. After partial evacuation the cathode is heated to a dull red heat to convert the carbonates to the oxides. The resulting carbon dioxide together with water vapor and other residual gases is driven off by this treatment. A test at this stage usually shows a relatively small electron emission. A so-called *activation* treatment is used to increase the emission. One such process consists in heating the cathode to about 1500°K. and at the same time impressing a voltage of about 150 volts on the anode through a suitable protective resistance. When the emission current builds up, the temperature of the cathode is reduced. The evacuation is completed at this stage, and the tube is sealed off. The cathode is ready for operation at a temperature of 900 to 1100°K.

The most important factors that affect the emission of the oxide-coated cathode are the core material, the coating materials, and the activation process employed. Core materials that have been used include nickel, tungsten, molybdenum, an alloy of

¹ *Ann. Physik*, **14**, 425, 1904.

platinum with 5 per cent nickel, and an alloy of cobalt, nickel, iron, and titanium called *Konel*. Both the particle size of the coating materials and the proportions of the constituents have a bearing on the emission. In general, the smaller particle sizes give more uniform emitting surfaces and increase the emission.¹ Different proportions of BaO and SrO have been investigated; about equal parts of the oxides give the best emissions.

There are two predominant factors in the activation process, (1) the heat-treatment and (2) the emission current drawn during activation. The experiments of Lowry² show that heat-treatment alone is sufficient to activate cathodes having Konel and platinum-iridium alloy cores; but other investigators find that the activation is considerably facilitated by drawing emission currents when the core material is a platinum alloy.

The phenomena associated with the emission from oxide-coated cathodes are exceedingly complex, and there is as yet no general agreement on explanations of some of the effects obtained. Lowry's experiments led him to conclude that the electrons were emitted from the core into the pores of the oxide and that they diffused through these pores, finally being reemitted from the outer surface of the oxide. A layer of Ba atoms³ or ions adsorbed at the core could conceivably account for such an emission. The activation by heat-treatment, according to this view, took place by the reduction of some of the oxide at the surface of the core metal, the barium being thus liberated. Later experiments⁴ including electron-microscope studies⁵ lead to the view that the electron emission is conditioned by free barium both on the outer surface of the oxide and dispersed through the oxide. The activation consists in the reduction of the barium at the core with subsequent dispersal outward by diffusion. The effect of emission current during activation is considered to arise from the electrolysis of the oxide with concomitant deposition of barium on the core and a resulting increase in the supply of

¹ BENJAMIN, M., R. J. HUCK, and R. O. JENKINS, *Phys. Soc. Proc. (London)*, **50**, 345, 1938.

² LOWRY, E. F., *Phys. Rev.*, **35**, 1367, 1930.

³ Both Ba and Sr may be present, but reference will be made only to Ba.

⁴ See general reference 6-1, Chap. IV, and articles by J. P. Blewett, *J. App. Phys.*, **10**, 668, 831, 1939.

⁵ HEINZE, W., and S. WAGNER, *Z. tech. Physik*, **17**, 645, 1936; **20**, 16, 1939.

barium available for outward diffusion. The main type of conduction through the coating is assumed to be electronic, only a small proportion being electrolytic. The principal effect of the activation is postulated to be the formation of a layer of adsorbed barium on the outer surface of the oxide crystals, which produces a strong lowering of the work function in much the same fashion as described for thorium on tungsten. Another effect is the dispersal of free barium through the oxide, which increases its conductivity.

A very low pressure of oxygen is sufficient to destroy the activation of an oxide-coated cathode. This is understandable, for the free barium would be converted to the oxide.

There are several difficulties connected with tests of Richardson's law as applied to coated cathodes. One is that there is a large uncertainty regarding the temperature scale for the different cathodes. Another is found in the failure of these cathodes to show a definite saturation current with increasing voltage. The results of different investigators can rarely be directly compared because there are wide differences in the preparation of the cathodes. The results presented by Arnold¹ in 1920 indicated the following as an average value of the emission constants for BaO and SrO mixtures on platinum-alloy cores as used by the Western Electric Company:

$$A = 0.47 \text{ amp./}(\text{cm.}^2)(\text{deg.}^2) \quad \phi_0 = 1.70 \text{ volts}$$

By 1929 the average constants for filaments of the same company were reported² to be as follows:

$$A = 0.01 \text{ amp./}(\text{cm.}^2)(\text{deg.}^2) \quad \phi_0 = 1.0 \text{ volt}$$

Values of work function below 1 volt have been reported, for example, by Thompson,³ who found

$$A = 0.03 \text{ amp./}(\text{cm.}^2)(\text{deg.}^2) \quad \phi_0 = 0.75 \text{ volt}$$

This cathode gives an emission of the order of 5 amp./sq. cm. at 1000°K.

c. Cathode Structures, Cathode Efficiency, and Effect of Operating Conditions.—The simplest type of cathode is the filamentary type, consisting of a straight filament, or sometimes of an inverted

¹ ARNOLD, H. D., *Phys. Rev.*, **16**, 70, 1920.

² DAVISSON, C., "International Critical Tables," vol. VI, p. 53, 1929.

³ THOMPSON, B. J., *Phys. Rev.*, **36**, 1415, 1801, 1930.

V or W or other similar forms. When it is desirable to use alternating current for heating the cathode, without producing a corresponding disturbance in the electron stream, the *unipotential* or *indirectly heated* cathode is employed. Figure 6-12a shows the essential features of two cathodes of this type. The tungsten heater wire is insulated from the electron-emitting nickel sleeve by an insulating coating on the wire. This is the most common cathode in use in modern receiving tubes, because of its freedom from electrical interference, its adaptability to various circuit arrangements, and its advantages over the filamentary type in the design of the tube itself.

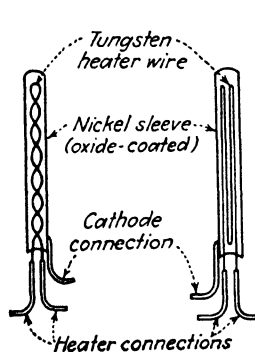


FIG. 6-12a.—Indirectly heated cathodes.

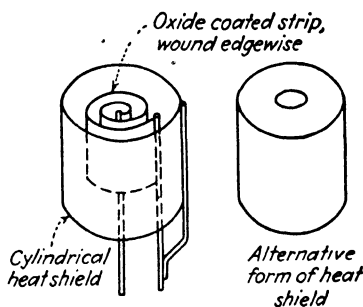


FIG. 6-12b.—Sketch showing the principle of the heat-shielded cathode useful in gaseous discharge tubes.

The cathode shown in Fig. 6-12b is fitted with a polished cylindrical heat shield. This reflects a portion of the heat radiated from the cathode spiral back toward the spiral again, thus conserving heat. The coated spiral itself is arranged in a form to conserve heat. Such cathodes are useful only in the gaseous- or vapor-discharge tubes, where the electrons can take part in the conduction even though emitted into slots or internally reentrant portions of the cathode surface. As will be clear after the treatment in Sec. 6-5, such portions would be prevented from emitting in a high-vacuum tube by the negative space charge that would collect and by the lack of an accelerating field at the cathode surface. Both the indirectly heated and the heat-shielded cathodes require a relatively long time to reach operating temperature. In large heat-shielded cathodes the heating time may be several minutes.

An important consideration in the application of hot cathodes to electronic devices is the cost of the power required to produce a desired electron emission. To facilitate such comparisons, it is convenient to use the so-called "cathode efficiency," defined as the ratio of the saturation emission to the cathode heating power. The units in the ratio should be stated, for example, in milliamperes per watt.

$$\text{Cathode efficiency} = \frac{\text{saturation emission current}}{\text{cathode heating power}} \quad (6-21)$$

Table 6-4 gives the range of cathode efficiencies found in commercial tubes. The values of the cathode efficiency would be

TABLE 6-4

Cathode	Approx. operating temperature, deg. K.	Cathode efficiency, ma./watts	Heating power, watts/sq. cm. of cathode surface
Pure tungsten*.....	2450-2500	2- 6	70
Thoriated tungsten*.....	1900	25- 80	20
Oxide-coated*†‡§.....	900-1100	40-2000	2.3-6

* General reference 6-7, Chap. VII.

† DUSHMAN, S., *Elec. Eng.*, **53**, 1054, 1934.

‡ THOMPSON, B. J., *Phys. Rev.*, **36**, 1415, 1801, 1930.

§ PATAI E., and G. FRANK, *Z. tech. Physik*, **16**, 254, 1935.

expected to vary especially widely for oxide-coated cathodes because of the variations in the surface properties previously discussed. These surfaces also show variations in their heat emissivities which also directly affect the cathode efficiency. The design of the cathode has much to do with the cathode efficiency; for example, the heat losses from the ends of the cathodes depend on the design. Higher cathode efficiencies can be obtained by raising the operating temperature, but at the expense of a shorter life. For example Davisson¹ gives a most probable life of 20,000 hr. for the Western Electric coated filament when operated at 1000°K. with a cathode efficiency of 24 ma./watt, whereas the life is 7,400 hr. when operated at a temperature of 1100°K. with a cathode efficiency of 53 ma./watt.

It has been mentioned already that oxygen adsorbed on tungsten filaments reduces the emission enormously. The active

¹ General reference 6-4, p. 427.

thoriated-tungsten and oxide filaments have a very strong affinity for oxygen, and these surfaces are deactivated, or "poisoned," by oxygen and water vapor and sometimes by carbon dioxide. Pressures of these gases of about 0.001 mm. Hg are sufficient to poison these cathodes, so that they must be operated in an extremely high vacuum. Inert gases, such as argon and neon, and mercury vapor have little effect on thermionic cathodes unless there is a discharge in the tube. Then the emitters may disintegrate or they may not, depending on the energy of the positive ions striking the cathode. Hull¹ has shown that thoriated-tungsten cathodes have fairly long lives in such gaseous-discharge tubes if the positive-ion energies are kept below

definite values, for example, 27 e.v. for Ne, 25 e.v. for A, and 22 e.v. for Hg.

In most vacuum tubes, there is at least a small production of positive ions by impact of the electrons with the neutral gas atoms. Tungsten filaments are best able to withstand the resulting positive-ion bombardment in high-voltage tubes. Thoriated-tungsten cathodes are used in intermediate-voltage power tubes, say up to 5,000-volt

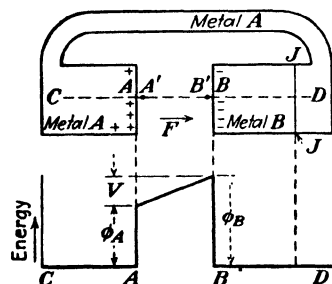


FIG. 6-13.—The relation of the work functions of two metals to their contact difference of potential.

tubes; and oxide-coated cathodes are used in receiving tubes, in hot-cathode gaseous light sources, and in gaseous rectifiers and thyratrons.

6-5. Contact Difference of Potential.—Let two metals A and B be joined at J as shown in Fig. 6-13. Experiments show that electrons cross the junction J in the direction from the metal having the lower work function into the metal having the higher work function until an equilibrium is set up. If metal A has the smaller work function ϕ_A , metal A will become positively charged, and B will become negatively charged.

The lower portion of Fig. 6-13 shows the changes in the potential energy of an electron carried along the line CABD from metal A to B and back into A. Starting with an electron at an arbitrary zero of energy at point A, the electron will gain an

¹ HULL, A. W., *Trans. A.I.E.E.*, 47, 753, 1928.

energy ϕ_A e.v. when it leaves A and moves out to point A' just outside the metal surface, say 10^{-3} cm. from the surface. As the electron moves from A' to B' , it will gain an energy V electron volts due to the electrostatic field intensity F between A' and B' . From B' to B the loss in potential energy is ϕ_B . If the change in energy of the electron from B through the junction to D is negligibly small, then we must have, since the energy at C and at D must be identical,

$$\phi_A + V - \phi_B = 0$$

or

$$V = \phi_B - \phi_A \quad (6-22)$$

This states that the *contact difference of potential between two metals equals the difference between their net work functions*. The above analysis neglects the Peltier e.m.f. at the junction J , which is usually small (1 per cent or less) compared with $\phi_B - \phi_A$, unless the two work functions are nearly equal.

Experiments designed to check Eq. (6-22) have not been altogether successful, although a fair agreement is found. The experiments are difficult because (1) it is hard to obtain a perfectly clean surface free of adsorbed gases, (2) different crystal faces of the same element may exhibit values of the work function differing by as much as 0.3 to 0.4 volt, and (3) the work function may vary with the temperature. The contact potentials of a series of 43 metals were measured with respect to a standard metal by Klein and Lange.¹ They made their measurements in an argon atmosphere at room temperature, and most of the metals were cleaned by grinding as the test proceeded. The contact potentials varied with time after each experiment started, but after 2 to 30 min. they approached a constant value. A few of their results for the work functions of metals deduced from these measurements and based on $\phi = 4.52$ e.v. for mercury are shown in Table 6-5.

A comparison of these values with those obtained by thermionic tests (Table 6-2) shows differences of 0.1 to 0.8 volt, some values being higher and some lower.

The relation between contact potential difference and work function has been applied² to the study of the changes in the work function of oxide-coated cathodes during activation. A practical

¹ KLEIN, O., and E. LANGE, *Z. Elektrochemie*, **44**, 542, 1938.

² HEINZE, W., and S. WAGENER, *Z. Physik*, **110**, 164, 1938.

result of the contact potential difference between the electrodes of diodes and triodes is its effect on the space-charge-limited current (see Sec. 6-7).

TABLE 6-5.—WORK FUNCTIONS BASED ON CONTACT POTENTIAL TESTS

Metal	Work Function, e.v.
Hg.....	4.52
Pt.....	4.52
Mo.....	4.48
Cu.....	4.46
W.....	4.38
Ni.....	4.32
Ta.....	3.96
Zn.....	3.66
Th.....	3.46
Ba.....	1.73
Na.....	1.6
K.....	1.6

6-6. The Schottky Effect. Field Emission. *a. The Schottky Effect.*—The slight increase in the saturation electron current as the anode voltage is increased (see *C* to *D*, Fig. 6-2) was first treated theoretically by Schottky.¹ His theory has been found to hold quite accurately for pure metal cathodes, but modifications of the theory are necessary for atomic-film and oxide-coated cathodes.

Schottky considered that in the saturation region the anode potential would cause an accelerating electric intensity at the cathode surface. This field intensity would result in a force on the electron in an outward direction from the metal surface, *i.e.*, opposed to the image force and surface polarization force. Thus, the force diagram in Fig. 6-8 must be modified to include the force caused by the electric field. If the metal surface is assumed to be smooth, then the intensity and hence the force will be a constant, as shown in the force diagram in Fig. 6-14. The original force curve *ABCD* is therefore changed in the presence of the electric field to the curve *FGH*. This new curve goes through zero at the point x_0 where the image force is just equal to the force caused by the external field. At all points to the right of x_0 the electron experiences a net force away from the

¹ *Physik. Z.*, **15**, 872, 1914.

surface. Hence, to become free, the electron needs only to reach the plane at $x = x_0$.

Let us consider what effect this state of affairs has on the energy diagram. The line OT shows the energy of the electron caused by the external field at any value of x . These values must be subtracted from the potential-energy curve for zero electric field to give the final curve LM . This curve has a

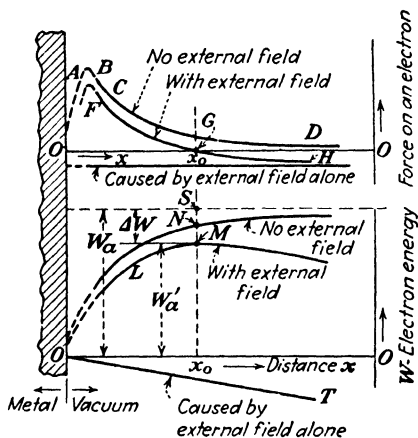


FIG. 6-14.—The effect of an external electric field on the force and energy curves.

maximum value at x_0 , for it is at this point that the net force curve passes through zero. The net effect, then, is to lower the outer work function from W_a to W'_a . The net work function will, of course, be reduced by the same amount.

To put these ideas in mathematical form: We may consider first that the distance x_0 is determined by the equality of the external force caused by the field and the image force, on the assumption that x_0 is not too small; thus, by using Eqs. (2-5) and (2-40),

$$Fq_e = \frac{q_e^2}{16\pi\epsilon x_0^2} \quad (6-23)$$

The reduction in the work function ΔW may be considered as arising in two portions, (1) the reduction in the work caused by the external field from $x = 0$ to $x = x_0$, which has the value $Fq_e x_0$; and (2) the reduction caused by the fact that the electron does not need to overcome the image forces from x_0 out to infinity. The first portion is represented by the vertical distance MN

in the energy diagram (Fig. 6-14), and the second by NS . Thus, the total reduction is

$$\Delta W = Fq_e x_0 + \int_{x_0}^{\infty} \frac{q_e^2}{16\pi\epsilon x^2} dx \quad (6-24)$$

Upon integrating, and eliminating x_0 by using Eq. (6-23), we find

$$\Delta W = \frac{q_e^{\frac{1}{2}} F^{\frac{1}{2}}}{(4\pi\epsilon)^{\frac{1}{2}}} \quad (6-25)$$

This reduction in the work function will have an effect on the emission current; using the form of Eq. (6-15), we find the emission current in an accelerating field to be

$$J_s = A_0 T^2 \exp \left[-\frac{q_e}{kT} \left(\phi_0 - \sqrt{\frac{q_e F}{4\pi\epsilon}} \right) \right] \quad (6-26)$$

or

$$J_s = A_0 T^2 \exp \left[-\frac{11,600}{T} (\phi_0 - 3.99 \times 10^{-5} \sqrt{F}) \right] \quad (6-27)$$

This equation can be tested experimentally by plotting the logarithm of the current density against the square root of the electric intensity at the cathode. A straight-line relation shows agreement with Eq. (6-27). Such straight lines have been obtained for tungsten filaments for fields up to 10^8 volts/m. when the wires have a small crystal structure and a very smooth surface. When the surface is slightly rough, there are deviations from the theoretical line caused by the variation of the intensity over the "hills" and "valleys" of the surface. Even sub-microscopic roughnesses may cause such deviations; for at a field of 10^8 volts/m. the value of x_0 , on the assumption of a smooth surface, is 1.9×10^{-6} mm.

The deviations in the curves for atomic-film and oxide-coated cathodes are greater than for pure metals. These large deviations have been quite satisfactorily explained¹ in terms of patches on the surface which have a work function different from that on the rest of the surface.

The most important application of the Schottky equation (6-27) is in the deduction of the true work functions of the metals from their Richardson lines. As already defined, the work function ϕ_0 refers to the work of extraction when the external

¹ See, for example, general reference 6-3.

field is zero. Equation (6-27) must be used to reduce the observed saturation current to the value that would be obtained with zero external field at the cathode surface. This can readily be done for pure metals, but the correction is quite uncertain for composite cathodes.

b. Field Emission.—When the field intensity is raised at the surface of a metallic cathode in a high vacuum to the order of 10^8 volts/m., an electron emission appears that is nearly independent of the temperature. It is observed at room temperature and hence is sometimes called *cold-cathode emission* but more generally *field emission*. At first, it was thought that the Schottky law would account for these currents; for if the field intensity is raised high enough, Eq. (6-27) shows that the current varies much more rapidly with the field intensity than with the temperature. However, the experiments of Millikan and Eyring¹ and of Ahearn² showed that cathodes of tungsten and molybdenum gave field currents which were independent of temperature to an accuracy of 5 per cent from 300 to 1100°K. Their results could be represented by an equation of the form

$$I = B \exp \left(- \frac{D}{F} \right) \quad (6-28)$$

where B and D are constants that depend on the kind of metal and its previous treatment.

The theory developed by Fowler and Nordheim,³ who used a wave-mechanical approach to the problem, gave as a result an equation similar to (6-28), as follows:

$$I = aF^2 \exp \left(- \frac{D}{F} \right) \quad (6-29)$$

Here, a and D are constants, and the exponential factor provides the major part of the variation with intensity. According to Fowler and Nordheim's theory, an electron that approaches an energy barrier, or hill, as shown by curve LM in Fig. 6-14, has a finite chance of escaping even though its energy is less than that of the top of the hill. The effect of an increase in external field is to lower the hill and make it narrower. The decrease in

¹ *Phys. Rev.*, **27**, 51, 1926.

² *Phys. Rev.*, **44**, 277, 1933.

³ *Proc. Roy. Soc. (London)*, **119**, 173, 1928.

height is most important in thermionic emission, and the decrease in width of the hill is considered to be sufficient to account for the field emission. The experiments cited above gave results that agreed in *form* with the theoretical results, but quantitative agreement was not obtained. The failure to obtain a quantitative check of the theory was ascribed to the influence of sub-microscopic roughnesses, perhaps in the form of protuberances or crystal edges in the form of ridges, on the actual electric intensity at the metal surface. Such roughnesses would need to increase the intensity by a factor of about 70, and investigators have not agreed whether such a large factor is reasonable.

The experiments are usually carried out with a very sharp point adjacent to a plane surface or with a wire of about 0.02 mm. diameter arranged down the center of a cylindrical anode. With unconditioned metals for the cathode, appreciable field-emission currents are obtained with fields of the order of 10^7 to 5×10^7 volts/m. Outgassing by high-temperature treatment increases the intensity at which the currents become measurable to the order of 8×10^7 volts/m. At these high fields the currents may follow a stable voltage-current curve for a certain range of voltage and then jump suddenly to a current as much as 10^4 or 10^7 times as high as before. This behavior has been called a breakdown¹ although the discharge after the breakdown follows a stable voltage-current curve. With carefully conditioned tungsten electrodes, fields as high as 5×10^8 volts/m. have been obtained.

Electron-microscope observations throw some light on the phenomenon. Using a magnification factor of about 200,000, Müller² obtained electron images of the field currents from a single crystal point. The emission was observed to come from minute regions of the order of 10^{-11} sq. cm. in area, and the "radius" of the extremity of the point was about 5×10^{-6} cm. The observed current density was about 10^6 amp./sq. cm.

The experiments of Beams³ indicated that a surface of impure mercury in a high vacuum (near the freezing point of Hg) gave a field emission starting at 3.5×10^7 volts/m. whereas a pure surface required an intensity of 1.8×10^7 volts/m.

¹ For example, by A. J. Ahearn, *Phys. Rev.*, **50**, 238, 1936.

² MÜLLER, E. W., *Z. Physik.*, **106**, 541, 1937.

³ *Phys. Rev.*, **44**, 803, 1933.

Special precautions must be taken to eliminate field emission from the electrodes of high-voltage high-vacuum tubes, such as kenetron rectifiers and Coolidge X-ray tubes, for the currents of 30 to 100 ma. that may build up would be destructive to the tube. The electron currents that result sometimes strike the glass walls, as manifested by the resulting fluorescent light, and actually give rise to destructive heating of the bulb. Field emission is also important in the theory of some types of arc (see Chap. XI).

6-7. Space-charge-limited or -controlled Electron Currents.—

The important applications of electron tubes depend on the control of the current through the tube by means of potentials impressed on one or more electrodes inside the tube. In the case of the two-element tube, or diode, the useful region of control is illustrated by the range from *A* to *B* in Fig. 6-2. This range has been discussed in a preliminary way in Sec. 6-1 where reference was made to the difficulty of the complete problem. Treatment of a simplified case will, however, give an insight into the role played by the space charge in the control of the electron current by the anode voltage. Therefore, consider Fig. 6-15 in which a parallel-plate anode and cathode, assumed infinite in extent, are separated by a distance *d*. It is further postulated that the electrons are just able to emerge from the cathode; *i.e.*, they emerge with a zero initial velocity. Consider now that the anode voltage E_1 is held constant and that the cathode temperature is slowly increased. While the cathode temperature is still low, the electron emission will be small. The electron current density will be very small, and the quantity of charge between the plates at any instant will not be sufficient to disturb the uniform field. The electrons emerging from the cathode will be in an accelerating field which will carry them to the anode. As the cathode temperature is increased, the rate of emission of electrons increases enormously, as shown in Fig. 6-4. This increased electron current will result in sufficient space charge to disturb the field, tending toward the potential distribution shown in Fig. 6-3, marked "with space charge." However, under the assumption of zero initial velocity of the electrons, the field could not be changed to the extent shown in Fig. 6-3, since the electrons would return to the cathode in the retarding field. Therefore, as the temperature is raised, the current will increase until the

electric intensity at the cathode approaches zero. A further increase in the supply of electrons cannot increase the space current, and it may be said that a condition of space-charge-limited emission has been reached. In what follows, it will be assumed that a very large supply of electrons is available at the cathode and, therefore, that the field intensity is practically zero at the cathode.

To express the effect of the space charge on the potential distribution between the plates, we require the one-dimensional form of Poisson's equation. This can be derived by applying Gauss's theorem to the elementary box cut from the right cylinder in Fig. 6-15 by the planes x and $x + dx$. The electric intensity everywhere on the narrow sides of the box is tangential to the surface; therefore, the integration in Gauss's law, *i.e.*, in

$$\int_{\text{closed surface}} F \cos (F, n) da = \frac{\Sigma Q}{\epsilon} \quad (2-32)$$

need be carried out only over the two ends of the box. If we consider an arbitrary electric intensity of magnitude $+F$ in the positive x -direction at

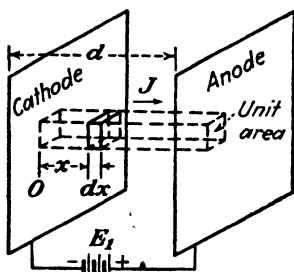


FIG. 6-15.—Parallel-plane diode.

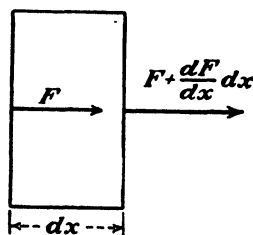


FIG. 6-16.—Side view of elementary box.

plane x , then the intensity at plane $x + dx$ is $F + \frac{dF}{dx} dx$ (Fig. 6-16). In accordance with convention, we take an outward component of intensity as positive so that Eq. (2-32) reduces to

$$-F + \left(F + \frac{dF}{dx} dx \right) = \frac{\rho dx}{\epsilon}$$

where ρ is the volume density of charge in the elementary volume, and the charge enclosed in the box is ρdx . This simplifies to $dF/dx = \rho/\epsilon$; and, by making use of the relation between intensity and potential [Eq. (2-22)], we may write the required form of Poisson's equation

$$\frac{d^2 E}{dx^2} = -\frac{\rho}{\epsilon} \quad (6-30)$$

Returning now to the space-charge problem, we let

ρ = volume density of charge at the plane x (Fig. 6-15).

E = potential at the plane x , referred to zero potential at the cathode.

q_e = value of the charge of the electron ($+1.6 \times 10^{-19}$ coulomb).

V = velocity of the electrons at the plane x .

J = current density, which is the same at any cross section in the steady state.

The emitted electrons are assumed to start from rest, so that the velocity of an electron at x is given by

$$V = \sqrt{\frac{2q_e E}{m}} \quad (3-7)$$

The moving electrons constitute a space charge of density ρ which is moving with a velocity V , parallel to the x -axis, so that the current density is given by

$$J = \rho V \quad (6-31)$$

The variables ρ and V can be eliminated by combining (6-31), (3-7), and (6-30), giving the following differential equation:

$$\frac{d^2 E}{dx^2} = -\frac{J}{\epsilon} \left(\frac{m}{2q_e} \right)^{\frac{1}{2}} E^{-\frac{1}{2}} \quad (6-32)$$

This may be integrated by multiplying through by $2 \frac{dE}{dx}$, with the following result:

$$\left(\frac{dE}{dx} \right)^2 = -\frac{4J}{\epsilon} \left(\frac{m}{2q_e} \right)^{\frac{1}{2}} E^{\frac{1}{2}} + K \quad (6-33)$$

The integration constant K is zero, for E is zero at the cathode, and we have postulated a great enough supply of electrons at the

cathode to make the intensity $\left(-\frac{dE}{dx} \right)$ zero. The variables in

(6-33) are separable, and a second integration yields

$$E = \left(\frac{-9J}{4\epsilon} \right)^{\frac{1}{2}} \left(\frac{m}{2q_e} \right)^{\frac{1}{2}} x^{\frac{4}{3}} \quad (6-34)$$

The integration constant is again zero, for the potential was taken to be zero at the cathode. Since the current flow is electronic, the current density J in the $+x$ -direction will be negative, thus

accounting for the minus sign in Eq. (6-34). This equation shows how the potential varies between the electrodes. The curve starts with a horizontal slope at the cathode and then rises as the four-thirds power of the distance.

A more useful result is obtained if Eq. (6-34) is solved for the current density in terms of the potential at the anode E_1 and the separation d . Thus, we find the magnitude of the current density to be given by the following relation:

$$J = \frac{4\epsilon}{9} \left(\frac{2q_e}{m} \right)^{\frac{1}{2}} \frac{E_1^{\frac{3}{2}}}{d^2} = 2.335 \times 10^{-6} \frac{E_1^{\frac{3}{2}}}{d^2} \quad (6-35)$$

Thus, the current varies as the three-halves power of the anode voltage and inversely as the square of the spacing. This shows that in order to drive electron currents across the space-charge region at reasonably low voltages the spacing between the cathode and anode must be small.

The foregoing theory applies with slight modification to the analogous case of positive-ion motions through a region of positive-ion space charge; in fact, the theory originally was developed for this case by Child.¹

If the effect of the initial velocities is taken into account, the resulting relation between the current and the anode voltage is modified somewhat. This problem has been attacked by Langmuir² and others. The initial velocities of the emitted electrons are assumed to be distributed according to Maxwell's law. Experimental tests on this point can readily be made, and the tests show that the assumption is in accord with the facts. The effect of the initial velocities is to cause a minimum in the potential-distribution curve as shown in Fig. 6-3. Experimental tests, as well as theoretical calculations, show that the negative minimum potential E_m and also its distance from the cathode x_m become smaller as the anode voltage is increased.

The expression given by Langmuir³ for the current-voltage relation is as follows:

$$J = \frac{2.335 \times 10^{-6} (E_1 - E_m)^{\frac{3}{2}}}{(d - x_m)^2} \left[1 + \frac{T^{\frac{1}{2}}}{40.3 (E_1 - E_m)^{\frac{1}{2}}} \right] \quad (6-36)$$

¹ *Phys. Rev.*, **32**, 492, 1911.

² LANGMUIR, I., *Phys. Rev.*, **21**, 419, 1923.

³ General reference 6-13, pp. 239ff.; general reference 6-3, Eqs. (86), (87), and (88).

This is the same as the approximate form (6-35) except that the potential and anode distance are measured from the potential minimum and the bracketed correction factor has been added. The magnitude of E_m is given in Eq. (6-41) below, and x_m has the algebraic value

$$x_m = 1.09 \times 10^{-6} T_1^{\frac{1}{2}} \zeta_1 J^{-\frac{1}{2}} \quad (6-37)$$

ζ_1 is a function of $\log (J_s/J)$; a few corresponding values are given in Table 6-6.

TABLE 6-6.—VALUES OF ζ_1

$\log \frac{J_s}{J}$	0	0.30	0.60	1.00	1.60	2.40	4.5	10	15	25
ζ_1	0	-0.979	-1.312	-1.60	-1.881	-2.117	-2.404	-2.544	-2.553	-2.554

The order of magnitude of E_m and x_m can be seen from the values in Table 6-7. These were calculated from Eqs. (6-41) and (6-37) for the parallel-plate case, assuming a separation

TABLE 6-7.—EFFECT OF INITIAL VELOCITIES ON SPACE-CHARGE-LIMITED CURRENTS

J , amp./sq. m.	E_m , volts	x_m , mm.	E_1 , volts	J/J_s	J [by Eq. (6-35)], amp./sq. m.
1,000	-0.235	0.0203	1,196	0.336	964
200	-0.58	0.057	399	0.0672	186
40	-0.94	0.146	130	0.01344	34.6
8	-1.27	0.347	40.3	0.00269	6.0
2	-1.58	0.697	13.1	0.000672	1.11

d of 0.01 *m*. The cathode is assumed to be tungsten at a temperature of 2500°K., for which the saturation-current density is 2,975 amp./sq. m. For convenience in making the calculations the value of J was set at five different arbitrary values. In addition to giving the values of E_m and x_m , the table also gives E_1 , calculated from Eq. (6-36), the ratio of the assumed current to the saturation current, and the current density calculated by the approximate equation (6-35). The table shows that the currents calculated by the more accurate methods are higher than those calculated by Eq. (6-35), the percentage difference being larger for smaller values of current. Therefore,

the current changes less rapidly with the voltage than predicted by the three-halves-power law, especially at low anode voltages. This effect increases with the temperature of the cathode.

The presence of the minimum in the potential curve gives rise to a retarding field from the surface of the cathode out to the point $x = x_m$. The electrons emitted from the cathode must have energies greater than just sufficient to become free of the metal, for they must expend energy to overcome the retarding field. The extra energy required is E_m e.v. per electron. Many of the electrons that become free of the metal will not have enough energy to reach x_m , and these will be turned back. Thus the concentration of electrons will be greatest at the metal surface and will decrease out to x_m . Beyond x_m the concentration decreases again, but here it is controlled by the space-charge conditions in the accelerating field. A precise formulation of the relative concentrations of electrons at the cathode and at $x = x_m$ can be obtained by applying Boltzmann's general relation¹ for the effect of a field of force on the concentration of gas particles, which may be written

$$\frac{N_1}{N_2} = e^{-\frac{W}{kT}} \quad (6-38)$$

In this equation, N_1 and N_2 are the concentrations at region 1 and region 2, W is the energy required to carry a particle from 2 to 1 against the forces of the field, and k and T have their usual significance. To apply this relation to the case in hand, let N_c be the electron concentration at the cathode and N_m the concentration at $x = x_m$; then,

$$\frac{N_m}{N_c} = e^{\frac{E_m q_e}{kT}} \quad (6-39)$$

if q_e represents the magnitude of the electronic charge and E_m is the algebraic value of the potential at x_m . Now, the current density of electrons leaving the cathode is proportional to N_c , and the current density passing x_m (hence, reaching the anode) is proportional to N_m ; therefore, Eq. (6-39) yields the expression

$$\frac{J}{J_c} = e^{\frac{E_m q_e}{kT}} \quad (6-40)$$

¹ LOEB, L. B., "Kinetic Theory of Gases," 2d ed., p. 92, McGraw-Hill Book Company, Inc., New York, 1934.

since the current leaving the cathode is the saturation current. Solving Eq. (6-40) for E_m , we find that

$$E_m = -\frac{kT}{q_e} \log \frac{J_s}{J} = \frac{-T}{5,041} \log_{10} \frac{J_s}{J} \quad (6-41)$$

An expression for the space-charge-limited current from a cylindrical cathode to a concentric cylindrical anode on the assumption of zero initial velocities was derived by Langmuir. The current for a length L of the cylinders is given by

$$I = \frac{8\pi\epsilon}{9} \sqrt{\frac{2q_e}{m}} \frac{LE_1^{\frac{3}{2}}}{r_a\beta^2} = 1.47 \times 10^{-5} \frac{LE_1^{\frac{3}{2}}}{r_a\beta^2} \quad (6-42)$$

The symbol r_a stands for the inner radius of the anode cylinder, and β^2 is a complicated function of the ratio of r_a to the cathode radius r_c . The values of β^2 may be read from a curve;¹ a few values are given in Table 6-8.

TABLE 6-8

r_a/r_c	1.0	2.0	5.0	10.0	20	40	100	200	400	∞
β^2	000	0.279	0.767	0.978	1.072	1.095	1.078	1.056	1.036	1.000

The equation for coaxial cylinders again gives a current proportional to the three-halves power of the cathode-to-anode voltage; furthermore, it can be shown that a three-halves-power law holds for any geometrical arrangement as long as the current is entirely controlled by space charge.

Equation (6-42) is important not only because it is used in the design of coaxial cylinder diodes but because it is used as the basis for the study of cylindrical tubes containing grids, as well. The equation is also used as an approximation in calculations pertaining to tubes having V-shaped filaments and an oval cylindrical anode.

The initial velocities of the electrons produce less deviation from Eq. (6-42) than those calculated for the parallel-plate case. However, a similar minimum in the potential curve appears in this case, also.

Other factors that produce deviations from the three-halves-power law are the contact potential between anode and cathode,

¹ See general reference 6-13, p. 247.

the effect of the potential fall along the filament caused by the heating current, and the voltage drop through the oxide coating of oxide cathode tubes.

6-8. Typical Applications of the High-vacuum Diode: Rectifiers; Voltmeters.

The principal useful feature of the diode is its ability to pass a current in one direction and to prevent passage of current in the opposite direction. This gives the diode its ability to rectify alternating currents. This feature of rectification was first used by J. A. Fleming in 1904 for the reception of radio signals, and it is still an important use of the tube.

The diode is also used for the rectification of alternating current for power purposes, although the gaseous diode, mercury-arc rectifier, and ignitron (see Chap. XI) are more important in this field.

The circuit shown in Fig. 6-17 may be used for obtaining a rectified direct current from the source of alternating voltage e . The circuit is a series circuit of three elements: the alternator, the load resistance R , and the diode. If we let e represent the instantaneous e.m.f. of the alternator, i the instantaneous value of the current, and e_d the instantaneous e.m.f. of the diode, then, using Kirchhoff's e.m.f. law, we can write

$$e - iR - e_d = 0$$

That is, the sum of the e.m.fs. taken around the circuit is zero.

The manner in which these e.m.fs. add to zero may be seen in Fig. 6-18. In this figure the alternator e.m.f. e has been represented as a sine function of the time. The diode conducts current during the half cycle while e is positive, since its anode is positive with respect to its cathode, but acts as an insulator while e is negative. During the nonconducting half cycle, the voltage across the diode rises to its highest value, marked E_{inv} in the figure. This maximum inverse peak voltage must be limited to a safe value in accordance with the rating of the tube. If the

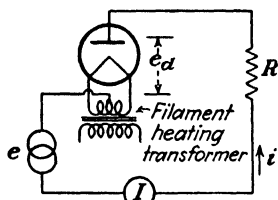


FIG. 6-17.—Half-wave rectifier circuit.

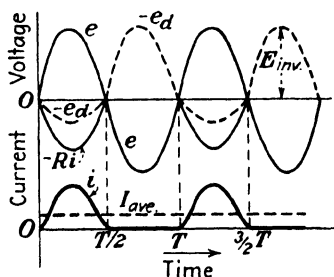


FIG. 6-18.—Current and voltage wave forms for a half-wave rectifier.

relation between e_d and i were linear, then the current would consist of half cycles of a sine curve. In general, however, the relation between e_d and i is nonlinear and is instead given approximately by the three-halves-power law:

$$i = ke_d^{\frac{3}{2}}$$

The nonlinear nature of this relation greatly increases the difficulty of calculations pertaining to a circuit containing a diode. On the other hand, this nonlinear characteristic has a useful application in one method of detection of radio signals.

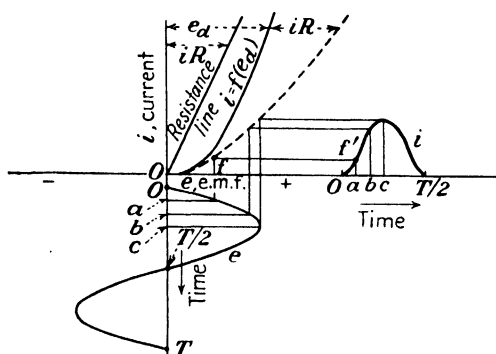


FIG. 6-19.—Construction of the current wave form for a half-wave rectifier.

The general problem presented by the diode rectifier circuit in Fig. 6-17 can be stated as follows: With a given alternating voltage e , a given load resistance R , and a known relation between the current i and the diode e.m.f. e_d , (a) what will be the average value of the current i , and (b) what is the wave form of the current i ? Because of the fact that the relation between i and e_d is often known only in an empirical form as a result of an experimental test, it follows that a graphical analysis is the most satisfactory approach to the problem. Let the experimental curve relating i and e_d be plotted as in Fig. 6-19. Then a straight line can be drawn to show the variation of the e.m.f. of the load resistance with the current. If corresponding abscissas of these two curves are added, the resulting curve will show the variation of the total back or counter e.m.f. with the current. The dotted curve in Fig. 6-19 gives this relationship. In order to find the current wave form, all that is necessary is to use the dotted curve to find values of current corresponding to the alternator voltage

at any instant. A convenient way of getting this result graphically is shown. The alternator voltage e is first plotted as a function of time, with the same e.m.f. scale as before but with a time scale in the downward direction from the origin. Then a time scale is laid out along the x -axis. The current curve is obtained by a point-by-point construction as follows: Take the instant a on the e.m.f. curve, for example. The current at this instant is found by projecting upward to the dotted curve, point f being thus obtained. This value of current is then projected horizontally to give the current at the instant a in the current curve, point f' . In this way the whole current wave form can be traced out. Further analysis could be made to find the average value of the current.

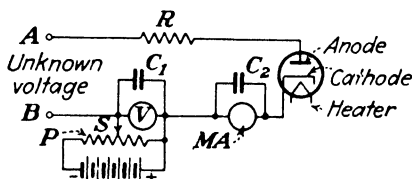


FIG. 6-20.—Circuit diagram of a peak voltmeter.

The method given above could be used to study such important features of a rectifier as its efficiency and the variation of the average voltage across the resistance as the resistance is varied. The method fails if the circuit contains inductive or capacitive circuit elements. The general problem is one of the solution of an electrical network containing a nonlinear element. The analysis of such circuits is outside the scope of this text.

The circuit of Fig. 6-17 can be used to measure alternating voltages. For this application the meter marked I would be a sensitive microammeter, and the resistance R would be made quite high. The advantage of this arrangement over an ordinary dynamometer or iron-vane type voltmeter is that the power consumed is relatively small and that higher frequencies can be applied to the device. However, the reading of the instrument tends to be proportional to the average value of the wave rather than to the r.m.s. value. This limitation can be reduced and the sensitivity can be increased by special methods described in texts on radio engineering.

The peak value of an alternating-voltage wave can be measured with the circuit depicted in Fig. 6-20. Let us assume that

if terminals A and B are short-circuited and if the potential divider P is adjusted so that V reads zero, then the microammeter will read zero. If, now, an alternating voltage is impressed across AB , the resulting rectified current will deflect the microammeter; but if the slider S is adjusted so that V equals the peak value of the voltage wave from A to B , the rectified current will fall to zero. This is called the *slide-back* voltmeter, because of the method of manipulation. C_1 and C_2 are high-frequency by-pass condensers. This voltmeter cannot be used with accuracy¹ when the peak voltage is below a few volts. At higher values of voltage the accuracy is better, but the method suffers the disadvantage that a battery voltage at least equal to the highest peak voltage to be measured must be available.

6-9. Grid Control of Electron Space-current: The High-vacuum Triode.—The fact that a third electrode in a high-vacuum thermionic tube will control the current was discovered by L. deForest in 1907. This electrode usually takes the form of a lattice or mesh of fine wires mounted between the cathode and anode and is called the grid. Such a tube is called a high-vacuum triode.

The internal action of the triode is similar to that of the diode, for in both tubes the current is controlled by the electrode voltages in accordance with the space-charge requirements. The space-charge conditions in the parallel-plane diode at any instant can be shown schematically as in Fig. 6-21*a*. The space charge in the region between the cathode and the plane of the potential minimum just outside the cathode (at x_m) has been omitted from the figure; therefore, all the space charges shown are in an accelerating field and are moving toward the anode. A simple consideration will show that this space charge must be the complementary charge of the anode charge; *i.e.*, the space charge between x_m and the anode must be equal in magnitude to the positive charge on the anode. This is obvious if one considers a Gaussian surface with one end in the anode and one end at the plane x_m where the intensity is zero and with the sides normal to the plane of the electrodes. An analysis of the results in Sec. 6-7 will show that when the anode voltage of the diode is increased there are two results which cooperate to increase the anode cur-

¹ See C. B. Aiken and L. C. Birdsall, *Trans. A.I.E.E.*, **57**, 171, 1938, for a study of the precision of the slide-back voltmeter.

rent: one is an increased velocity of the electrons at any plane, and the other is an increased space-charge density at the same plane.

Now, let us insert a grid of fine parallel wires between the cathode and anode as in Fig. 6-21b, and let the grid be at cathode potential. The grid will partly shield the cathode-to-grid region from the influence of the anode, thus lowering the intensity in this region and raising the intensity between the grid and anode. The cathode-to-grid region may be considered in terms of an equivalent diode with a spacing equal to the cathode-grid spacing which produces field intensities approximately equal to those

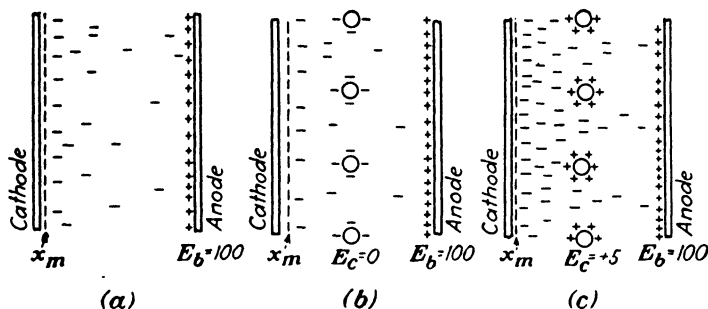


Fig. 6-21.—Charge distributions in the diode and triode. E_c = cathode-to-grid potential. E_b = cathode-to-anode potential.

actually present in this region. An analysis of the equations for the diode (Sec. 6-7) will show that if the intensities are decreased then the accompanying space-charge-controlled current is decreased, also. As a result, we conclude that the effect of the grid will be to decrease the current through the triode below what it was for the diode in Fig. 6-21a. The only effect of the increased intensities in the grid-to-anode region is to accelerate the electron stream and thereby reduce the space-charge concentration in the grid-to-anode space.

Another way of viewing the effect of the grid is to consider the induced charge on the grid. This negative induced charge on the grid (Fig. 6-21b) combined with the negative space charge forms the complementary charge of that on the anode. Thus the space charge is less dense than it was in the diode, and in the important region between the cathode and the grid the intensity is less; as a result, the current carried by the electrons is much reduced but is not eliminated entirely.

When the grid is charged positively, as shown in Fig. 6-21c, the space-charge density is increased, and the electron current leaving the cathode is increased. The electron stream is collected partly by the grid and partly by the anode, giving rise to a grid current as well as an anode current.

A small change in grid potential, as from zero to 5 volts (see Fig. 6-21b and c) can produce a relatively large change in the current. A theoretical analysis¹ shows that if the greater portion of the space charge is close to the cathode, as usually is the case, then the effects of the grid and anode potentials are linearly superposable, with the result that the electron current leaving the cathode can be given by the functional relation

$$I = f(\mu E_c + E_b) \quad (6-43)$$

where E_c is the cathode-to-grid potential, E_b is the cathode-to-anode (plate) potential, and μ is a coefficient of the tube called the *amplification factor*. An empirical form of the function is as follows:

$$I = B(\mu E_c + E_b)^n \quad (6-44)$$

where B and n are constants for a given tube. There is theoretical support for a value of $\frac{3}{2}$ for n , derived by an extension of the theory for the diode. However, experimentally determined values of n vary from 1.2 to 2. The value $n = 1$ is often employed in Eq. (6-44) to yield a form amenable to mathematical calculation.

An interesting result of Eq. (6-44) is that current continues to flow through the tube for negative grid voltages as long as μE_c is less than E_b , *i.e.*, as long as the quantity in the parenthesis remains positive. Interpreted physically, this means that, if the grid is not too far negative, it will allow electrons to pass through between its meshes. The field of the anode reaches through between the grid wires to produce an accelerating field near the cathode, and electrons are guided toward the anode by this field. The regions on the anode where the electrons strike can be nicely illustrated by means of a demonstration triode² which has a coating of fluorescent material on its anode. As

¹ VAN DER BIJL, H. J., "The Thermionic Vacuum Tube and Its Applications," McGraw-Hill Book Company, Inc., New York, 1920; see general references 6-2 and 6-7 for detailed accounts.

² Westinghouse type WL-787.

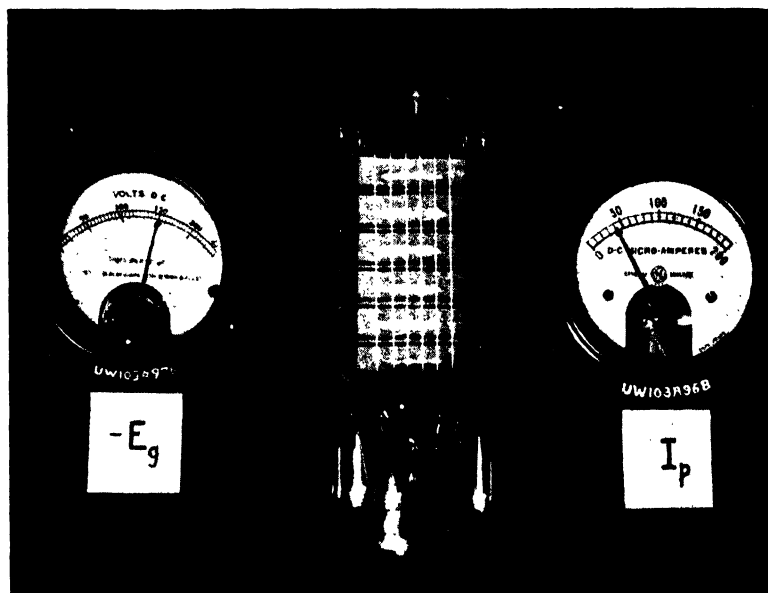


FIG. 6-22a.—A demonstration triode with -150 volts on the grid.

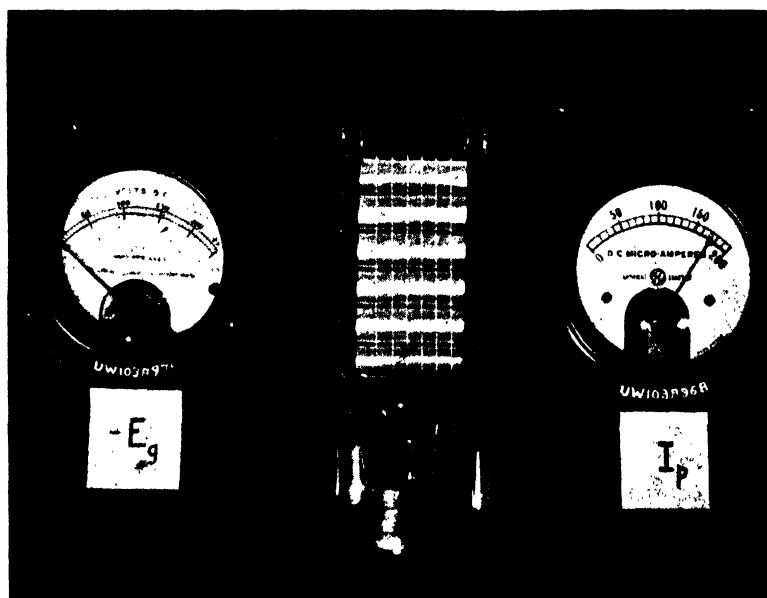


FIG. 6-22b.—The demonstration triode with zero volts on the grid.

shown in Fig. 6-22 the filament consists of a row of six vertical wires, and the grid consists of a rectangle fitted with horizontal cross wires. The anode is in the rear. In Fig. 6-22a the grid is at a negative potential of 150 volts, but the anode voltage of 400 volts is sufficient to develop a plate current of 40 μ a. The resulting fluorescence on the plate is rather faint, but the repelling action of the grid on the electron stream can be clearly seen. In Fig. 6-22b the grid voltage has been reduced to zero so that the anode current is much increased. This condition corresponds roughly to that in Fig. 6-21b. Electrons passing near the grid are still repelled so that the regions where the electrons strike the anode consist of bands which extend the full width of the anode. When the grid potential is made positive, the bands become wider and finally overlap. A detailed study of the paths of the electrons within the tube would involve the construction of field maps showing the intensity lines and equipotential surfaces.

The amplification factor μ in Eq. (6-44) gives a measure of the relative effect of a voltage on the grid and a voltage on the plate on the electron current through the tube. The grid is close to the cathode and has a greater effect on the dense space-charge region near the cathode than does the plate, and so a change of a few volts in the grid potential produces the same change in the current as a much larger change in the plate potential. Therefore, μ is greater than unity and in commercial triodes ranges from 3 to 100. Changes in the structure of a triode that increase its μ are the following: (1) an increase in the grid-wire diameter, (2) a closer spacing of the grid wires, and (3) a decrease in the spacing from cathode to grid.

In certain special forms of grid, cathode, and plate structures the value of μ can be calculated from theoretical formulas.¹ These formulas are based on the following considerations: In the absence of space charge, a grid voltage E_g acting in conjunction with a plate voltage E_b will produce a charge Q on the cathode given by the equation

$$Q = C_{kg}E_g + C_{kp}E_b \quad (6-45)$$

where the capacitance coefficients C_{kg} and C_{kp} are dependent only on the geometry of the tube. To apply this equation, we postulate that the plate and grid voltages have the same relative effect on the space-charge-limited current as they do on the cathode charge in the absence of space charge.

¹ General reference 6-2, p. 178.

This assumption will seem more plausible when one reflects that in one sense the space charge replaces the charge Q which would be present on the cathode in the absence of space charge. For if we think of a gradual introduction of the space charge into the field, we see that the cathode charge starts at the value Q and finally drops to zero when the current is completely space-charge-controlled. The cathode charge is actually not quite zero but would be zero if an equivalent cathode were imagined located at the plane of minimum potential at x_m . To find the relative effect of the grid and anode potentials from Eq. (6-45), we divide by C_{kp} .

$$Q = C_{kp} \left(\frac{C_{kg}}{C_{kp}} E_c + E_b \right) \quad (6-46)$$

The coefficient of E_c is identified with the μ of the triode.

$$\mu = \frac{C_{kg}}{C_{kp}} \quad (6-47)$$

The capacitance coefficients can be calculated from the geometry of the tube in certain cases, and formulas for μ obtained in this way agree satisfactorily with the measured value of μ . An expression¹ that applies to a concentric-cylinder filamentary triode with a helical grid follows:

$$\mu = \frac{2\pi N r_1 \log \frac{r_2}{r_1} - \log \cosh \pi N D}{\log \coth \pi N D} \quad (6-48)$$

where N is the number of grid wires per unit length of grid, r_1 is the radius of the grid helix, r_2 is the inner radius of the anode, and D is the diameter of the grid wires. When $\pi N D$ is less than 0.4, i.e., when the diameter of a grid wire is less than 12.5 per cent of the grid-wire spacing (center to center), then the formula reduces to the following approximate form:

$$\mu \approx \frac{2\pi N r_1 \log \left(\frac{r_2}{r_1} \right)}{\log \left(\frac{1}{\pi N D} \right)} \quad (6-49)$$

This is accurate to 5 per cent when μ is greater than 2.

In circuits employing the triode, the utilization device is placed in the plate connection so that it carries the plate current. In these applications, therefore, the functional relation of most interest is that between the plate current I_b and the control voltages, grid voltage E_c and plate voltage E_b . If the electron current collected by the grid is negligible, then Eq. (6-44) can be applied.

$$I_b = B(\mu E_c + E_b)^n \quad (6-50)$$

¹ VOGDES, F. B., and F. R. ELDER, *Phys. Rev.*, **24**, 683, 1924.

The assumption regarding negligible grid current will be satisfied in the important region of negative grid voltages where the grid potential is more negative than about 1 volt with respect to the cathode. As already mentioned, the more accurate forms of Eq. (6-50) often lead to mathematical difficulties in circuit problems. Therefore, various approximate methods and methods of graphical analysis are often employed. The usual procedure is to represent the relation between plate current and grid and plate voltages by a series of characteristic curves in which the relation between two of the variables is shown in a curve and the third is used as an adjustable parameter.

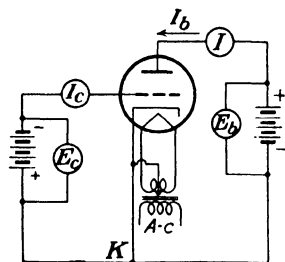


FIG. 6-23.—Circuit for the determination of triode characteristics.

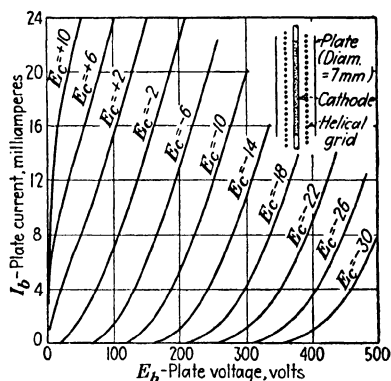


FIG. 6-24.—Plate characteristics of a triode (type 76). E_c = cathode-to-grid voltage.

The essential features of a circuit which may be used to obtain the data for the characteristic curves are shown in Fig. 6-23. The tube shown has an indirectly heated cathode. The grid and plate voltages should be made controllable, for example, by means of potential-dividing rheostats. When the triode has a filamentary cathode heated with direct current, it is customary to measure grid and plate voltages with respect to the negative end of the filament; thus, in Fig. 6-23 the connection at K would be made to the negative end of the filament.

The relation between plate current and plate voltage, called the *plate characteristic*, is shown in Fig. 6-24. These curves refer to a small receiving type of triode having a μ of about 14 and a cylindrical structure as shown in the figure. A study of Fig. 6-24 shows that the characteristic curves are quite similar for the different fixed grid voltages except in the region of positive grid voltages.

The relation between plate current and grid voltage with the plate voltage as the parameter is shown in Fig. 6-25a. The technical name of this type of curve is the *transfer characteristic* between the grid and plate. This name applies, in general, to any curve between the current to one electrode and the variation in the potential of another electrode while all remaining electrodes are maintained at specified constant potentials.

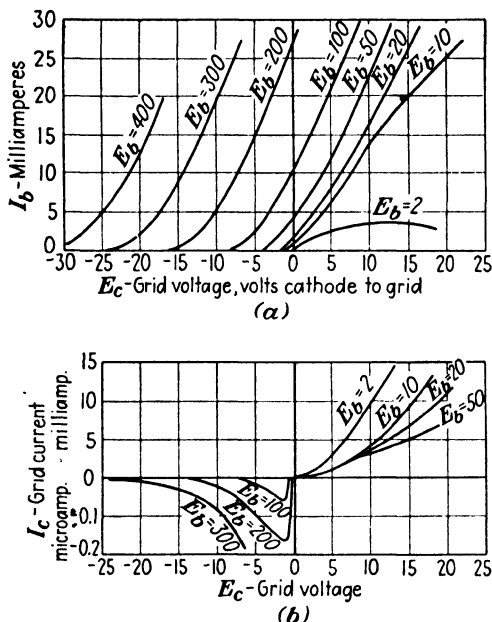


FIG. 6-25.—(a) Transfer characteristics, and (b) grid characteristics of a triode (type 76). E_b = cathode-to-anode voltage.

The grid-current characteristics are shown in Fig. 6-25b, in which positive grid currents are plotted in *milliamperes* and negative grid currents in *microamperes*. In the region of positive grid potentials the grid currents are largely dependent on the electron paths in the electric field between the electrodes. For convenience, we may divide the positive grid potentials into two classes, (1) those less than the plate potential, and (2) those greater than the plate potential. For potentials in class 1 the electrons are in an accelerating field from cathode to grid and again from the grid toward the plate. The electrons are accelerated along the lines of electric intensity as explained in Chap. III;

but the electrons tend to leave the convex sides of curved intensity lines because of their inertia. This fact, coupled with the small projected area of the grid relative to the plate area, accounts for a fairly small grid current relative to plate current. For the tube in Fig. 6-25, for example, the grid collects less than one-third as many electrons as the plate when both have the same positive potential and collects a lesser proportion when the grid potential is not so high as that of the plate. When the grid potential is higher than that of the plate (class 2), the electrons are in a retarding field from grid to plate; *i.e.*, they are accelerated toward the grid in all parts of the tube. Thus, electrons that miss the grid wires in their flight away from the cathode may or may not strike the plate. Of course, the energy corresponding to their component of velocity normal to the plate is the only part that is useful in overcoming the retarding field. The less energetic electrons will describe parabolalike curves which miss the plate and finally terminate on the grid. Therefore, as the grid potential exceeds the plate potential the grid current will increase at an accelerating rate while at the same time the plate current increases at a lesser rate or even begins to fall off. This is evident in the I_b curves for $E_b = 10$ volts and for $E_b = 2$ volts in Fig. 6-25a. When the electrode voltages exceed 10 or 20 volts, a complicating phenomenon arises caused by the *secondary emission* of electrons from the grid and plate by virtue of the electron impacts. This type of emission will be discussed in the following section.

The most important region of operation of the type of triode under discussion is with negative grid potentials. The curves in Fig. 6-25b show that in this region the grid currents are very small, at most a few tenths of a microampere. Therefore, a variation of the grid voltage produced, say by a radio signal, is accompanied by only a negligible power input to the grid of the tube. At the same time the grid causes variations in the plate current which may serve to control relatively large amounts of power. It is this property of controlling the current in the plate circuit by means of a negligible power input to the grid which has made possible such extensive application of the grid-controlled tube.

Although the negative grid currents are small, they are objectionable if the grid circuit contains a high resistance, say over

1.5 megohms for the small tube under discussion. The negative grid currents can originate in three ways, (1) by positive ions flowing toward the grid, (2) by leakage over the insulation, and (3) by electron emission from the grid wires. An appreciable amount of gas remains in the tube even in the best possible vacuum, so that the electrons will occasionally encounter a gas molecule with sufficient energy to ionize the molecule. The resulting positive ion either will be collected by the negative grid or will travel to the cathode. An increase in plate current will cause an increase in the positive-ion current. Since the positive-ion current decreases as the vacuum is improved, a test of the negative grid current is often employed to indicate the vacuum conditions in a tube. Although electron emission from the grid is usually negligible, it must be considered when very small grid currents are sought. This emission can arise by the following processes: photoelectric effect caused by the light from the cathode or by soft X rays from the anode, thermionic emission, and electron release by positive-ion bombardment of the grid. The most important causes of negative grid current usually are the positive-ion current and the leakage current. In Fig. 6-25*b* the curves show that the current is mostly positive-ion current, since the grid current goes nearly to zero when the plate current goes to zero. Special tubes, called electrometer tubes, are designed to have very low negative grid currents of the order of 10^{-14} to 10^{-15} amp. These tubes can be used in suitable circuits to measure currents as small as 10^{-17} amp.

In this section, we have seen that the effect of the grid on the electron stream is much larger than that of the plate. The amplification factor μ has been interpreted in two ways, (1) in terms of the geometry of the tube [Eq. (6-47)] and (2) in terms of the current-voltage relations [Eq. (6-44)]. The forms of the experimentally determined curves between the plate current and the control voltages, as well as between the grid current and the grid and plate voltages, have been discussed. Methods of applying these curves to the solution of typical tube circuits will be presented in Sec. 6-11.

6-10. Secondary-electron Emission Caused by Electron Impact.—Although the emission of secondary electrons from metals by electron impact was discovered about 1900, the effect began to receive more attention when it was found to produce

undesirable variations in the characteristics and behavior of thermionic tubes. More recently, the effect has been put to work in devices used in television; examples are the electron multiplier and the iconoscope.

A simple experiment on a thermionic triode will demonstrate the phenomenon, although this test is not suitable for quantitative measurements. Let the grid of the triode shown in Fig. 6-26 be fixed at a high positive potential, and let the cathode temperature be so low that the grid collects a small temperature-limited emission current. Then the plate current is observed as the plate potential is varied from zero up to that of the grid. As shown schematically in the circuit diagram in Fig. 6-26, a

primary electron P may release a secondary electron S at the plate, and S will be accelerated toward the grid, thus adding to the grid current. At the same time, the effect of the emission of the secondary electron S is to *reduce* the net plate current I_p . This effect begins to be noticeable in the plate-current curve at A . Some tubes yield a curve like $OABCD$. The zero value of plate current at B in this curve

is explained by an emission of one secondary electron for every primary electron that strikes the plate. The plate current is negative when more than one secondary electron is emitted for each primary electron. When the plate potential approaches the grid potential, the field strength between the grid and plate decreases, until finally there is no force attracting the secondary electrons toward the grid. Then the plate current rises rapidly as shown at D , since the secondary electrons fall back into the plate. Sometimes the curve obtained is similar to $OAFD$. In this case the ratio between secondary electrons and primary electrons does not reach unity; hence, the plate current remains positive.

In typical tests designed to produce quantitative results an electron beam passes through one or more apertures and then falls on the surface to be investigated. This surface is sur-

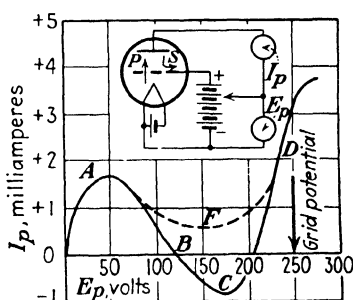


FIG. 6-26.—Curves showing the effect of secondary emission on the plate current.

rounded by a nearly closed conductor which collects all the secondary electrons and any of the primary electrons that may be reflected. This arrangement may be used to measure the distribution of the energies of the secondary electrons, and a modified arrangement is used to study the angles at which the secondary electrons are emitted. Their energy distribution is very nearly Maxwellian, having a broad maximum at a few electron volts energy, say 2 to 10 e.v. An interesting feature of these results is the fact that the energy distribution is independent of the energy of the incident primary beam, at least

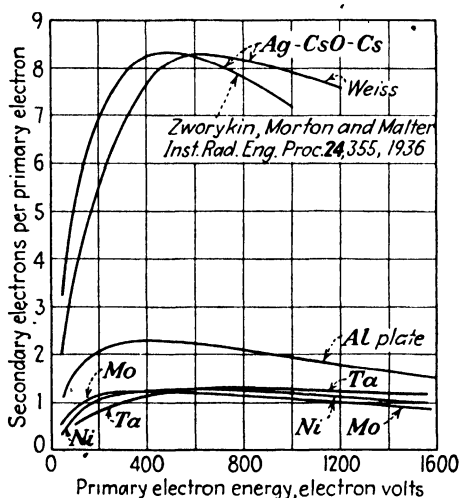


FIG. 6-27.—Secondary electron yields.

from 40 to 1,000 e.v. primary-electron energy, except for a small group of electrons that have energies close to the primary-electron energy. These are interpreted to be primary electrons that have been reflected without loss of energy. Actually, it is only on the basis of the energies that the secondary electrons can be distinguished from reflected primary electrons. The angular distribution tests showed that most secondary electrons were emitted normal to the surface and that the numbers at other angles decreased according to the cosine law. Both the energy distribution and angular distribution were found to be nearly independent of the angle of incidence of the primary electrons.

Of particular interest are the results on the number of secondary electrons emitted per incident primary electron. This number is

influenced by the energy of the primary electrons, the kind of metal, and the condition of its surface. Thus, outgassing the metal may decrease the secondary emission by a factor of 2 or 3. Results showing the effect of primary-electron energy on the secondary-electron yield are shown in Fig. 6-27 for a few representative surfaces.¹ A typical curve goes through a broad maximum at 300 to 1,000 volts. The curves for several pure metals not shown in the figure are similar to those given for Mo, Ni, and Ta. Thus, the curve for tungsten is close to that for tantalum; the curve for copper is slightly above that for nickel, and that for iron is slightly lower. The composite silver-cesium oxide-cesium surfaces give the greatest secondary-electron yields of any yet discovered.

No definite relation between the secondary emissions of pure metals and their thermionic work functions has been found. In fact, recent workers² find that the secondary emission from low-work-function metals such as Li, Mg, and Ba depends to a greater extent on the condition of the surface than was formerly believed. Though experiments on composite surfaces are not in general agreement, they show that a relatively large change, say 2 volts, in the work function of a given type of surface, for example, W-Th, has little effect (10 to 30 per cent) on the secondary emission.

The experimental results described in the foregoing paragraphs, together with the results of experiments on the passage of electrons through thin metal foils, have led to the following interpretation of the phenomenon: The incident electron penetrates into the metal crystal lattice and may impart some of its energy to one of the bound electrons of the metal. This electron has some chance of starting off in the direction of the metal surface; and if it is not absorbed or deflected from its path by the metal atoms, it will emerge as a secondary electron. According to this view the secondary electrons originate at a depth of 3 to 15 atomic diameters for 500-volt primary electrons. When the primary electrons have higher energies, they penetrate more deeply, but they transfer most energy to the bound electrons near the end of their path. Fewer secondary electrons emerge

¹ See summary by R. Kollath, *Physik. Z.*, **38**, 202, 1937.

² BRUINING, H., and J. H. DE BOER, *Physica*, **5**, 17, 1938. BRUINING, H., *Physica*, **5**, 901, 913, 1938. KOLLATH, R., *Ann. Physik*, **33**, 285, 1938.

because there is more chance for absorption and scattering before they reach the surface.

The grid of a transmitting triode may reach positive potentials of 200 volts or more. Most metals will emit 1.1 to 1.4 secondary electrons per primary electron when bombarded with 200-volt electrons. If the plate of the triode is at a higher potential than the grid, then the plate collects the secondary electrons from the grid, and the grid current will be reversed. This action is exactly the same as that discussed in connection with Fig. 6-26 except that the roles of plate and grid are interchanged. This phenomenon can cause *blocking* of the transmitter tube if the circuit is such that a negative grid current tends to keep the grid potential positive. Another troublesome effect caused by secondary-electron current in a screen-grid tube will be discussed in the following section.

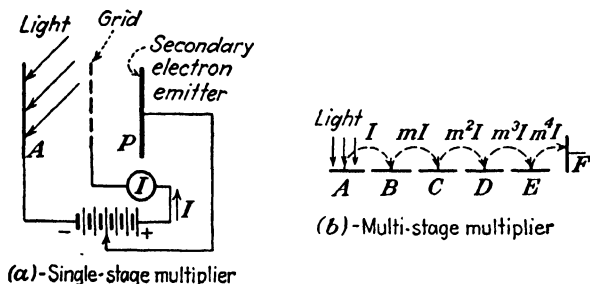


FIG. 6-28.—Electron multipliers.

The fact that one primary electron can yield more than one secondary electron makes it possible to design a secondary emission amplifier, or *electron multiplier*.¹ Figure 6-28a shows one scheme of amplification applied to photoelectric currents. The photoelectric surface A emits electrons, which are accelerated toward the grid by the field. A fraction of these primary electrons are collected by the grid, but the major portion pass through and strike the plate P. If each electron striking P knocks out eight or nine secondary electrons that return to the grid, then the current I may be as much as six times the original photoelectric current. The amplification obtainable can be increased by the scheme shown in Fig. 6-28b. In this device the photoelectrons emitted from surface A are accelerated toward the secondary emitter B by an accelerating voltage impressed on electrode B. The secondary electrons from B are again accelerated toward C by a voltage impressed between B and C, etc., until the current at collector F is m^4 times the original current, where m is the multiplying factor of each secondary emitter. The electrons must pass from surface to surface in well-defined beams formed by methods already discussed in Chap. III. Multipliers having an over-all amplification as high as 10^6 to 10^7 have been

¹ ZWORYKIN, V. K., G. A. MORTON, and L. MALTER, *Proc. I.R.E.*, **24**, 351, 1936. J. R. PIERCE, *Bell Lab. Rec.*, **16**, 305, 1938.

built on this principle. The output currents are limited to a few milliamperes, however, on account of the disturbing effect of space charge.

Secondary electrons are also emitted from *insulating surfaces* bombarded with electrons; but experiments on this effect are difficult, and little is known about the details of the process. When the number of secondary electrons emitted is *less* than the number of primary electrons, the insulating surface will acquire a negative charge. In consequence, the number of primary electrons reaching the surface is decreased, and positive ions in the discharge will be attracted toward the surface. When an equilibrium exists, the positive-ion current plus the secondary-electron current will just equal the primary-electron current. A different situation occurs when the number of secondary electrons exceeds the number of impinging electrons. Then the insulating surface becomes positively charged so that the primary electrons are accelerated toward the surface. As a result, there is extra heating where the primary electrons strike. This heating has been observed on the glass walls of high-voltage high-vacuum tubes and is occasionally so intense that a hole is melted in the glass. The spots where the electrons strike are indicated by the blue or green fluorescence on the glass.

6-11. Multigrid Thermionic Tubes; Applications of Grid-controlled Tubes.—The triode suffers the disadvantage, especially in radio-frequency circuits, that the capacitance between the grid and plate promotes undesirable oscillations in the associated circuits. This capacitance is reduced to a very small value in the screen-grid tube, or *tetrode*, by interposing a screen grid g_2 between the control grid g_1 and the plate P as shown in Fig. 6-29a.

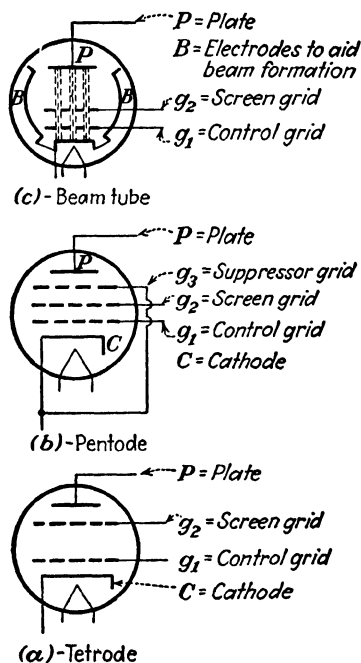


FIG. 6-29.—Schematic diagrams of multigrid high-vacuum tubes.

In some designs the screen grid has a fine mesh and nearly surrounds the plate. The screen grid is held at a positive potential of the order of one-third of the plate potential, with the result that the action of the electrons in the region from cathode to screen is the same as that in the cathode-to-plate space of the triode. Part of the electrons are collected by the screen, but ordinarily the larger portion travel to the plate. The course of the plate-current-plate-voltage curve for this tube is shown in curve *OCDS* in Fig. 6-30. The peculiar dip in the curve is explained by the effect of the secondary electrons released from the plate. When the plate potential is less than the screen-grid potential E_s , the secondary electrons emitted by the plate are

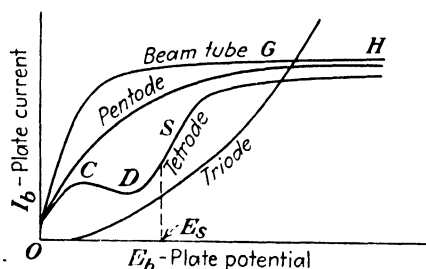


FIG. 6-30.—Plate characteristics of multigrid tubes. All electrode potentials are fixed except the plate potential.

collected by the screen grid and thus add to the screen current and reduce the plate current (see region *D*, Fig. 6-30). When the plate potential exceeds the screen potential by a few volts, the secondary electrons return to the plate; consequently, the plate current rises suddenly as in the range *DS*. At still higher plate potentials the plate current is nearly constant, because the screen grid shields the space-charge region near the cathode from the influence of the plate.

The falling portion *CD* corresponds to a negative variational resistance. This is called the *dynatron* region of the tetrode characteristic. It can be used in the generation of oscillations in much the same fashion as by the use of the falling characteristic of the arc (see Sec. 11-6).

The distorted region *CDS* in the tetrode characteristic limits the range of linear amplification and is undesirable. The distorting effect of the secondary electrons is reduced in the *pentode*, or five-electrode tube, by inserting a third grid g_3 , called the

suppressor grid, as shown in Fig. 6-29b. This grid is connected to the cathode as shown. As a result, a secondary electron emitted from the plate is in a field that forces the electron back to the plate. The plate characteristic that results is shown in Fig. 6-30. The suppressor grid has a rather wide mesh so that the main electron stream passes between the grid wires; furthermore, there is not too much retarding action, for the potential in the space between the suppressor-grid wires is not far below the plate potential. A potential drop of 10 or 15 volts from plate to the plane of the suppressor is sufficient to drive nearly all the secondary electrons back to the plate.

An action similar to that of the suppressor grid is obtained in the *beam tube* (Fig. 6-29c) by the effect of a dense space charge between its screen grid and its plate. By careful design the electrons are constrained to travel between the meshes of the control grid and the screen grid in well-defined beams. The space-charge effect is enhanced by the greater distance from screen to plate than in the ordinary tetrode. When the plate potential is less than the screen potential, there is enough space charge between the screen and the plate to produce a potential minimum in this region, and thereby the field distribution needed to drive secondary electrons back into the plate is produced. The resulting plate characteristic is shown in Fig. 6-30.

Multigrid tubes are used in much the same way as triodes; *i.e.*, the current through a load resistance or impedance in the plate circuit is controlled by the variations of the control-grid potential. In these tubes the amplification factor μ is defined to be the ratio of an increment in plate voltage Δe_b to the change in the grid voltage in the opposite direction Δe_c required to keep the plate current constant, while all other electrode voltages are held constant. Thus, the amplification factor can be expressed as follows:¹

$$\mu = \frac{\Delta e_b}{\Delta e_c} \left(i_b \text{ constant; all other elec-} \right) \quad (6-51)$$

trode voltages constant

For reasonably high plate potentials, *i.e.*, in region *GH* in Fig.

¹ Equation (6-51) is consistent with Eq. (6-50); for if I_b in Eq. (6-50) is constant, then $\mu E_c + E_b$ is constant. In terms of corresponding changes in E_c and E_b then, $\mu \Delta E_c$ must equal ΔE_b to keep I_b constant. Therefore, $\mu = \Delta E_b / \Delta E_c$ for constant I_b . Equation (6-51) is identical with this result but is written in terms of the instantaneous voltages e_b and e_c .

6-30, the plate current of the tetrode, pentode, or beam tube changes only slightly as the plate potential changes. The control grid, on the other hand, can control the plate current in the same proportion as in a triode. Consequently, the amplification factor of these tubes is generally high (order of 100 to 1,500).

The tubes discussed thus far have control grids with equal spacings between the grid wires so that the grid controls the electron flow equally all along the cathode. If, however, the grid-wire spacing is wide along part of the cathode and narrow

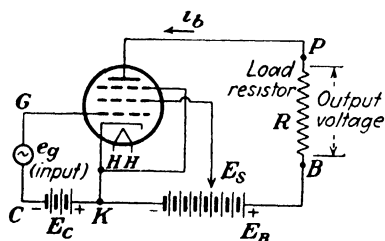


FIG. 6-31.—Fundamental amplifier circuit.

along the rest, the control is nonuniform. As a result, a much higher negative grid voltage is needed to stop all the electrons. Moreover, the amplification factor of the tube will not be a definite quantity but will depend on the magnitude of the plate current. Such tubes, called

variable- μ , or *supercontrol*, tubes, are used where the adjustable μ is desirable as in volume control in radio receivers.

Important applications of grid-controlled tubes are made in amplifiers as used in radio receivers and long telephone-line repeaters; in oscillators as found in radio transmitters and in carrier-current telephony; in modulators as used in radio telephony; in demodulators, or detectors, as found in radio receivers, and in electronic instruments such as the vacuum-tube voltmeter. Since the treatment of these diverse applications is beyond the scope of this book, we shall content ourselves with the graphical analysis and an introductory mathematical analysis of a simple amplifier circuit and with a qualitative examination of other typical applications.

A typical amplifier circuit using a pentode tube is shown in Fig. 6-31. A triode, tetrode, or beam tube can be used in place of the pentode, and the method of analysis to be described applies equally well when any of these tubes is used. The input signal voltage e_g is considered to be a sinusoidal alternating-voltage source of negligible internal impedance. The battery E_c supplies the negative grid bias, and battery E_B supplies both the screen-grid and plate circuits. The load resistance R is placed in the

plate circuit, so that variations in the plate current i_b caused by changes in the control-grid voltage will produce corresponding variations in voltage across R . When this output voltage is utilized to actuate the grid of another amplifier tube or to operate a high impedance device like a cathode-ray oscillograph, then the amplifier is called a voltage amplifier. In other cases the load resistance R represents the utilization device itself, such as a loud-speaker; and in this event the purpose of the amplifier is to develop as much alternating power in R as is consistent with limitations of available signal voltage and permissible distortion.

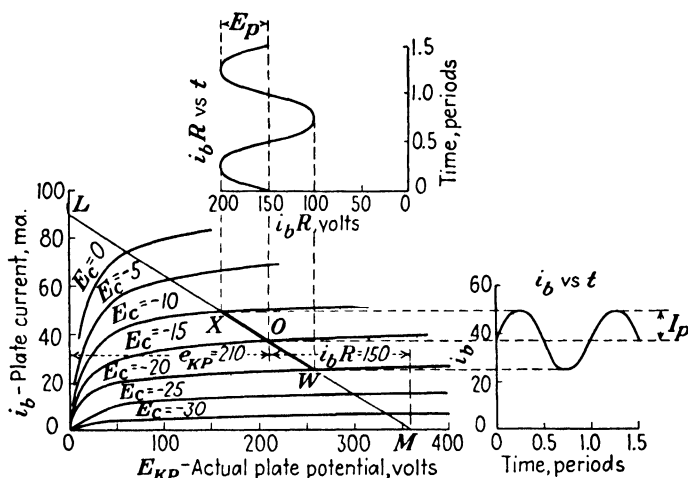


FIG. 6-32.—Graphical analysis of the amplifier circuit in Fig. 6-31.

The amplifying circuit in Fig. 6-31 can be analyzed graphically with the aid of the family of plate characteristics for the tube in question. Characteristics of this type are given in the tube manufacturers' handbooks,¹ and a typical set for a pentode is shown in Fig. 6-32. It will be convenient to begin the analysis with a discussion of the variation of the actual plate potential, e_{KP} , with the instantaneous plate current i_b . Applying Kirchhoff's e.m.f. law to the plate circuit, we can write

$$e_{KP} = E_B - i_b R \quad (6-52)$$

This is the equation of the straight line ML in Fig. 6-32. We note

¹ For example, "Receiving Tube Manual," RCA Mfg. Co., Harrison, N.J. issued annually.

that point M must be at the plate battery voltage E_B . The slope of the line can be found by locating point L on the current axis, as follows: At point L the actual plate potential is zero, or, in effect, the plate battery voltage is all consumed across the resistance R . Consequently, the ordinate of L is E_B/R . The line ML , called the *load line*,¹ can be used to aid the solution of a variety of problems relating to the amplifier. For example, suppose E_B is 360 volts, R is 4,000 ohms, and E_c is -15 volt, and we desire to find the value of the plate current i_b when the signal voltage is zero. The load line in Fig. 6-32 corresponds to the given values of E_B and R ; therefore, the point O , where the plate characteristic for $E_c = -15$ volt intersects the load line, gives the required plate current for zero signal voltage. At this point the actual plate potential is 210 volts, the voltage drop across R is 150 volts, and the plate current is 37.5 ma. Point O is called the *operating point* of the amplifier.

An important question, namely, What is the shape of the output voltage wave corresponding to a given input voltage?, can now be answered. For if the control-grid voltage is varied, the plate current and potential must vary along the line ML in order to satisfy Kirchhoff's e.m.f. law. Consequently, the plate current at any value of instantaneous grid potential is found at the intersection of the plate characteristic for this grid potential with the load line. For example, when the grid potential of the amplifier changes to -20 volt, the plate current will be 25 ma., and the voltage e_{KP} will be 260 volts, as shown by the intersection at W in Fig. 6-32. Likewise, if the grid potential changes to -10 volt, then the plate current is 49 ma. and the voltage e_{KP} will be 160 volts as shown at point X . A study of Fig. 6-32 will show that for small signals a condition obtains in which the plate current is a linear function of the grid voltage. This condition holds as long as the range of operation along the line covers regions where equal increments in the grid voltage give equidistant intersection points of plate characteristic with the load line. This criterion is approximately satisfied by the range WX in Fig. 6-32. Therefore, a sinusoidal grid voltage having a peak value of 5 volts will produce a plate-current wave as shown at the right in Fig. 6-32. The pulsations in the current are sinusoidal and vary between the ordinates corresponding to

¹ Sometimes called the resistance line (see Sec. 10-8).

points X and W with the frequency of the signal. There will be a corresponding variation of voltage across load resistance R . This voltage wave is illustrated in the top part of Fig. 6-32, where the peak of the wave is located by the vertical projection from point X and the minimum is similarly located from point W . The remainder of the wave is constructed as a sinusoid. The waves of total grid voltage, plate current, and voltage across the load are drawn on a common time axis in Fig. 6-33. The conventions as to polarity and symbols used in Fig. 6-33 are

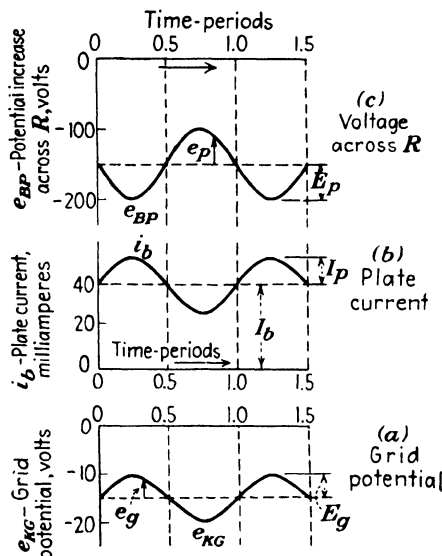


FIG. 6-33.—Potential and current waves in an amplifier circuit.

those customarily employed. The meaning of the double-subscript symbols for voltages is evident from an examination of Fig. 6-31. Then the signal voltage can be written as follows:

$$e_g = E_g \sin \omega t \quad (6-53)$$

The useful part of the voltage wave across the load resistor is the alternating component, which is given by

$$e_p = E_p \sin (\omega t - \pi) \quad (6-54)$$

A study of Figs. 6-32 and 6-33 will clarify the reason for the 180-deg. phase angle between e_g and e_p . The ratio of the output voltage to the signal voltage is called the *voltage amplification* of the amplifier; and, in terms of the symbols adopted above, we

have

$$\text{Voltage amplification} = \frac{E_p}{E_g} \quad (6-55)$$

The voltage amplification in the amplifier discussed above is $\frac{50}{5}$ or 10.*

When the load resistance is a device like a loud-speaker, we are interested in the alternating power developed therein. If we limit the discussion to the small-signal case, we see from Fig. 6-33*b* that the plate current can be expressed as

$$i_t = I_b + I_p \sin \omega t. \quad (6-56)$$

The total instantaneous power has the value

$$p = i_t^2 R = I_b^2 R + 2I_b I_p R \sin \omega t + I_p^2 R \sin^2 \omega t \quad (6-57)$$

The average power P will be the average of the instantaneous power over a complete cycle. The second term on the right in Eq. (6-57) drops out in the averaging process, and as a result we find

$$P = I_b^2 R + \frac{I_p^2 R}{2} \quad (6-58)$$

The first term on the right may be associated with the average plate current I_b , and the second term with the alternating component, so that the second term represents the useful power output, P_0 . Therefore

$$P_0 = \frac{I_p^2 R}{2} \quad (6-59)$$

Since the power input to the grid of the amplifier is very small, it is possible to realize a large gain in the available power at the signal frequency.

It is instructive to study the way in which the power supplied by the plate battery is consumed in the amplifier circuit. Obviously, part of this power is consumed in the resistance R as given by Eq. (6-58), and part is consumed in the tube itself. By appealing to the law of conservation of energy, we can say that the power P_t consumed in the tube is the difference between the power supplied by the battery and that consumed in the resistance R , thus:

* The amplification factor μ of the tube employed in the analysis above is approximately 180, and it should be noted that the voltage amplification developed in the amplifier is considerably below this figure.

$$P_t = E_b I_b - I_b^2 R - \frac{I_p^2 R}{2} \quad (6-60)$$

This power is expended in accelerating the electrons toward the plate but is finally liberated as heat developed in the plate by the electron impacts on the surface of the plate. It is interesting to note that, as the output power $\frac{1}{2}I_p^2 R$ increases, the power supplied by the battery stays constant, whereas the power expended in the tube decreases. In one sense, the conversion of battery power into the alternating form occurs at the expense of the power going to the tube itself.

The foregoing discussion has been based on the assumption of small signals, which give a nearly linear relation between plate current and grid voltage. The following analysis will show how the behavior of the amplifier may be predicted when the signal voltage is increased: Let the signal voltage wave have a peak of 15 volts and the other conditions remain the same. The operation will take place along the same load line ML as before but will extend to the intersections with the plate characteristics for grid voltages $E_c = 0$ and $E_c = -30$ volt. These intersections obviously locate the positive and negative peaks of the plate-current wave. A third point is known; for when $e_g = 0$, the operation is at point O . The intermediate intersection points are used to locate additional points on the plate-current wave in the following way: The intermediate grid voltages at the intersections in the graph are used to find the corresponding instants of time on the signal wave. Thus since the signal voltage wave is $e_g = 15 \sin \omega t$, we can easily find that for $E_c = -10$ volt, for example, e_g is +5 volts; and this value occurs at a time of 0.054

TABLE 6-9

E_c , volts	e_g , volts	Corresponding times, periods	Corresponding plate current, ma.
0	+15	0.25	75
- 5	+10	0.116, 0.384	63
-10	+ 5	0.054, 0.446	49
-15	0	0, 0.5, 1.0	37
-20	- 5	0.554, 0.946	25
-25	-10	0.616, 0.884	14.5
-30	-15	0.75	6

period after $t = 0$, and again at 0.446 period. Other values are given in column 3 of Table 6-9. The table also shows the corresponding plate currents read from Fig. 6-32.

The plate current is now known at enough instants on the cycle so that a wave can be plotted, as shown in Fig. 6-34. This wave is seen to be slightly distorted, the bottom half being slightly flat-topped, whereas the upper half wave is nearly sinusoidal. In this example, the distortion has the effect of raising the average value of the plate current above the value at point O , although in other cases the opposite result is found.

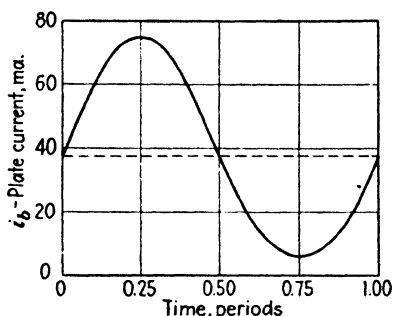


FIG. 6-34.—Plate current wave form.

The graphical analysis described above can be used to study the effect of changes in the plate battery voltage, load resistance, or operating point on the output of the amplifier. In the following discussion an introduction to the analytical treatment of amplifiers is presented. For more extended treatments the reader is referred to texts on radio engineering and on vacuum-tube applications.¹

The instantaneous plate current is a function of the grid potential e_g and the plate potential e_b and so may be expressed in the general form,

$$i_b = f(e_g, e_b) \quad (6-61)$$

An approximate expression for the plate current is obtained by expanding i_b in a Taylor's series about the operating point and neglecting second order and higher terms, thus:

$$i_b = I_b + \frac{\partial i_b}{\partial e_g} e_g + \frac{\partial i_b}{\partial e_b} e_p \quad (6-62)$$

where I_b is the plate current at the operating point, and e_g and e_p are the variable components of grid and plate potential. Usually we are concerned with the variable component of plate current,

¹ TERMAN, F. E., "Radio Engineering," 2d ed., McGraw-Hill Book Company, Inc., New York, 1937. REICH, H. J., "Theory and Application of Electron Tubes," McGraw-Hill Book Company, Inc., New York, 1939. General reference 6-2.

i_p , which, from Eq. (6-62), is given by

$$i_p = \frac{\partial i_b}{\partial e_c} e_c + \frac{\partial i_b}{\partial e_b} e_p \quad (6-63)$$

In this equation, $\partial i_b / \partial e_c$ and $\partial i_b / \partial e_b$ are the partial derivatives evaluated at the operating point.

A consideration of Eq. (6-63) shows that these partial derivatives have the dimensions of conductance and, in fact, are defined to be the *tube conductances*, as follows:

$$\text{Grid-plate transconductance} = G_{pg} = \frac{\partial i_b}{\partial e_c} \quad (6-64)$$

$$\text{Plate conductance} = G_p = \frac{\partial i_b}{\partial e_b} \quad (6-65)$$

Then Eq. (6-63) may be written as follows:

$$i_p = G_{pg} e_c + G_p e_p \quad (6-66)$$

This equation is useful in the analysis of tube circuits in which operation is limited to nearly linear portions of the tube characteristics. In such cases the equation is a good approximation, and the neglect of the higher order terms in Taylor's series does not introduce serious error.

As indicated in Eq. (6-64), the transconductance G_{pg} can be evaluated by taking the slope of the i_b vs. e_c curve at the operating point. For example, consider the tube of Fig. 6-25, and choose as an operating point $E_b = 200$ and $E_c = -10$ volt. Then the value of G_{pg} is found as the slope of the transfer characteristic at this point, expressed in amperes per volt or mhos. At the point chosen this slope is $1,520 \times 10^{-6}$ mho. Similarly, the plate conductance G_p is the slope of the plate characteristic at $E_b = 200$ and $E_c = -10$ in Fig. 6-24 and is found to be 107×10^{-6} mho. In the neighborhood of this operating point, therefore, the variable component of plate current is given by

$$i_p = 0.00152e_c + 0.000107e_p$$

The application of Eq. (6-66) to the simple amplifier in Fig. 6-31 will lead to a formula for the voltage amplification developed by the amplifier. The voltage e_p arises by virtue of the current i_p through the load resistance R , thus:

$$e_p = -Ri_p \quad (6-67)$$

Combining Eqs. (6-66) and (6-67) and solving for i_p , we find

$$i_p = \frac{G_{pg}e_g}{1 + G_p R} \quad (6-68)$$

To find the output voltage, we multiply Eq. (6-68) by R , with the result that

$$e_p = \frac{RG_{pg}e_g}{1 + G_p R} \quad (6-69)$$

and the voltage amplification e_p/e_g is obtained,

$$\frac{e_p}{e_g} = \frac{RG_{pg}}{1 + G_p R} \quad (6-70)$$

As an example of the application of this equation, we use the tube constants evaluated above, choose a load resistance of 40,000 ohms, and calculate a voltage-amplification ratio of 11.5.

To show the relation that exists between the tube conductances and its amplification factor, we express the variable part of the plate current as follows [see Eqs. (6-50) and (6-61)]:

$$i_b = f(\mu e_c + e_b) \quad (6-71)$$

Upon calculating the tube conductances by the application of Eqs. (6-64) and (6-65), we obtain

$$G_{pg} = \frac{\partial i_b}{\partial e_c} = \mu f'(\mu e_c + e_b) \quad (6-72)$$

and

$$G_p = \frac{\partial i_b}{\partial e_b} = f'(\mu e_c + e_b) \quad (6-73)$$

Dividing (6-72) by (6-73), we find that

$$\frac{G_{pg}}{G_p} = \mu \quad (6-74)$$

Thus, the tube coefficients G_{pg} , G_p , and μ are related by Eq. (6-74); and if two of the coefficients are known, the third can be calculated.

An alternative form of Eq. (6-70) is obtained if we divide numerator and denominator by G_p and replace G_{pg}/G_p by μ and $1/G_p$ by R_p .

$$\frac{e_p}{e_g} = \frac{R\mu}{R_p + R} \quad (6-75)$$

In this equation, R_p is termed the *plate resistance*, and, as indicated above, it is defined as the reciprocal of the plate conductance,

$$R_p = \frac{1}{G_p} \quad (6-76)$$

Thus we see that the voltage amplification of the amplifier can be expressed in terms of the tube coefficients G_{p0} and G_p or in terms of μ and R_p .

If we divide the numerator and denominator of Eq. (6-75) by R , we get

$$\frac{e_p}{e_g} = \frac{\mu}{1 + (R_p/R)} \quad (6-77)$$

From this equation, we see that when R is large compared with R_p the voltage amplification of the amplifier approaches the amplification factor of the tube. However, the maximum permissible value of R is often limited by practical considerations.

A second important application of grid-controlled tubes is in the production of modulated waves. Reference to Fig. 6-35 will aid the following qualitative discussion of the principle of operation of a typical amplitude modulator. The high-frequency, or *carrier*, voltage e_c is impressed in the grid circuit. A low-frequency voltage, which we designate as the voice-frequency voltage e_v , is impressed in the plate circuit. The grid bias is made sufficiently negative so that the plate current flows only during part of the time. The plate current is determined by the simultaneous action of waves e_c and e_v (see Fig. 6-36), and the resulting wave of i_b is shown at c . When the circuit is properly adjusted, the output current i_o will have the desired modulated form shown at d .

A third important application of grid-controlled tubes is in oscillator circuits that take energy from a d.c. source and deliver alternating power. In this respect the oscillator is very similar to the amplifier, but it is different in that the grid alternating voltage is obtained from the oscillating circuit. This can be accomplished in a variety of ways; the principal features of a common

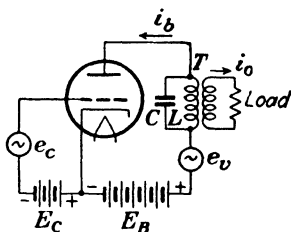


FIG. 6-35.—Circuit diagram of a modulated amplifier.

method are shown in Fig. 6-37a. When this circuit is properly adjusted, the tube maintains a steady alternating current around the loop CL_bL_c . Under ordinary conditions the circulating cur-

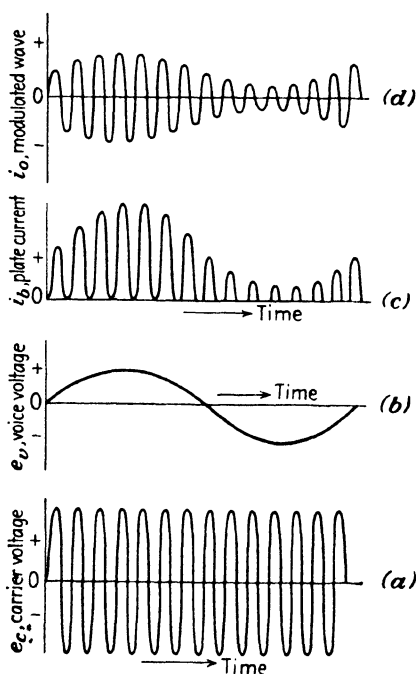


FIG. 6-36.—Wave forms in a modulator circuit.

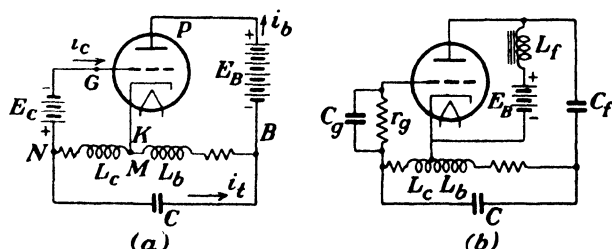


FIG. 6-37.—The Hartley oscillator. (a) The circuit in principle and (b) practical modification.

rent in this loop is large compared with the plate current i_b , so that the alternating voltages from K to B and from K to N are nearly sinusoidal. The waves of plate potential e_{KP} and grid potential e_{xK} are shown in Fig. 6-38a and b for a typical condition of

operation. These two potentials serve to determine the instantaneous plate current i_b , and its variation is shown at c . Plate current flows for only a fraction of the period because the grid bias is large. The grid potential becomes positive at its peak, and so grid current flows as shown at d . The dip in the top of the plate-current wave arises because the instantaneous plate

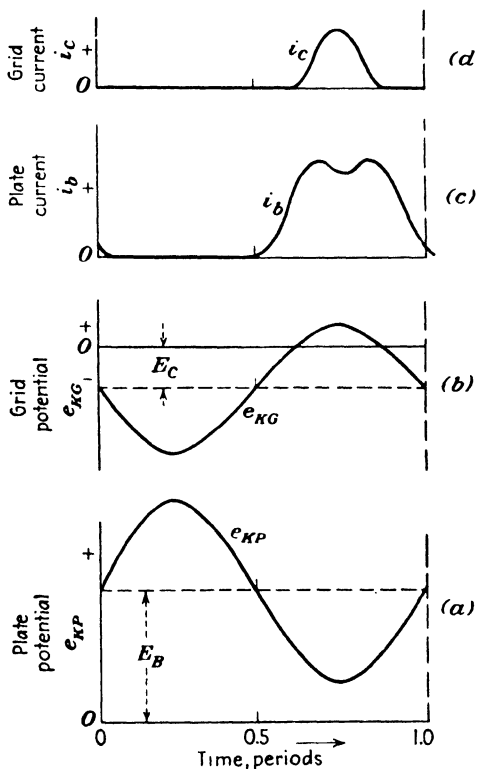


FIG. 6-38.—Wave forms in an oscillator circuit.

potential drops so low at this instant that the grid potential is nearly the same as the plate potential. The resulting grid current serves to reduce the electron stream going to the plate.

A brief qualitative discussion of the manner in which the tube maintains the oscillations will be of interest. A study of the circuit and of Fig. 6-38 will show that the pulse of plate current comes at a time when the voltage of the oscillating-circuit condenser C is near its maximum. The pulse of current may be

thought of as sending a pulse of charge through the tube which divides at point K in the circuit, with part of the pulse passing through L_c and going to charge condenser C to a still higher voltage, while the remainder returns through L_b to the battery E_B . The added energy in condenser C corresponding to the added charge is dissipated in maintaining the oscillation during the next cycle. In the steady state, then, the battery E_B delivers energy to condenser C which in turn delivers the energy to the oscillating circuit.

The oscillator of Fig. 6-37a will also function, but with reduced efficiency, when the bias voltage E_c is so small that plate current flows during the whole cycle.

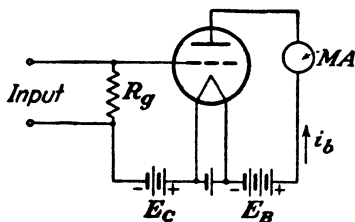


FIG. 6-39.—Triode voltmeter circuit.

When the amplitude of the oscillations is small, the behavior of the oscillator in Fig. 6-37a may be predicted by mathematical analysis; however, the practical circuits such as those shown in Fig. 6-37b are more often designed by semiempirical methods.

Triodes are frequently used in voltmeter circuits. The triode voltmeter suffers some of the disadvantages of the diode voltmeter, but it has the advantage of a very high input impedance. The principle of a common form of triode voltmeter is shown in Fig. 6-39. The grid bias E_c is large enough so that the plate current is low on the curved part of the E_c - I_b characteristic. The effect of an input voltage is to raise the average plate current read by the d.c. meter MA . The readings of the milliammeter can be standardized in terms of known input voltages. The input resistance R_g can be made very high, say 3 megohms, or can be dispensed with entirely if there is a d.c. path through the source of input voltage. This voltmeter, therefore, draws a negligible power from the test circuit.

6-12. Photoelectric Electron Emission.—The observations of Hertz in 1887 on the effect that ultraviolet light produced on the sparking distance between gaps in his famous experiments on electric waves led up to the experiments of Hallwachs. Hallwachs, in 1888, published an account of the first direct experiments which showed that negative electricity was given off from metals when ultraviolet light shines on the metal surface. Elster

and Geitel subsequently (1889) prepared surfaces of sodium and potassium amalgam that were sensitive to ordinary visible radiation.

An experimental arrangement for studying the photoelectric effect is shown in Fig. 6-40. Experiments show that the photoelectric effect consists in the emission of electrons from the sensitive surface under the influence of the light or radiation striking the surface. In the apparatus in Fig. 6-40 the emitted electrons are collected by the positively charged grid or wire collector and cause a current in the external circuit. The magnitude of this current is found to depend on several conditions of the experiment including the following: (1) the magnitude and polarity of the collector voltage E , (2) the intensity and wave-length characteristics of the incident radiation, (3) the material and condition of the sensitive surface, and (4) the type and pressure of the residual gas in the tube. After a brief account of the experimental results,¹ we shall discuss their theoretical implications.

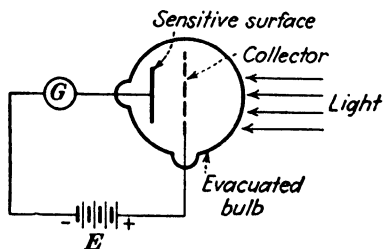


FIG. 6-40.—Photoelectric tube and circuit.

a. Experimental Observations.—If the collector-to-cathode voltage impressed on a highly evacuated photoelectric tube such as that shown in Fig. 6-40 is varied, a current-voltage curve results that is similar to curve *A* in Fig. 6-41. A considerably different type of curve is obtained (curve *B*, Fig. 6-41) when the collector is made in the shape of a sphere with a small hole that allows light to enter and strike a small centrally located cathode. The curve shows that in this case practically all the electrons reach the anode even with zero potential difference.

Tests of the relation of the photoelectric current to the intensity of the incident light show that under suitable testing conditions the current is directly proportional to the light intensity. This linear relation holds over remarkably wide ranges of current, especially with the high-vacuum photoelectric tube.

Suppose that radiation including a narrow band of wave lengths but of constant intensity is allowed to fall on a photo-

¹ For extended discussions, see general references 6-8 and 6-4.

electric tube and the resulting current is observed for various parts of the spectrum. The results of this experiment are plotted in the form of a *spectral distribution curve*, which relates

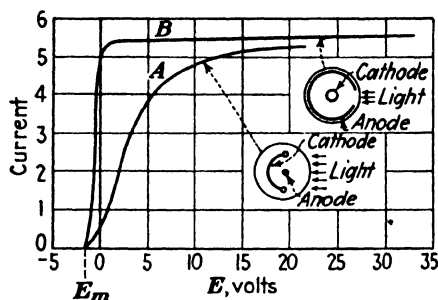


FIG. 6-41.—Current-voltage curves with constant illumination. E represents the cathode-to-anode voltage corrected for contact potential difference.

the current per unit power in the incident radiation to the wavelength or the frequency.¹

Typical spectral distribution curves obtained with ordinary light at normal incidence are shown in Fig. 6-42. When the

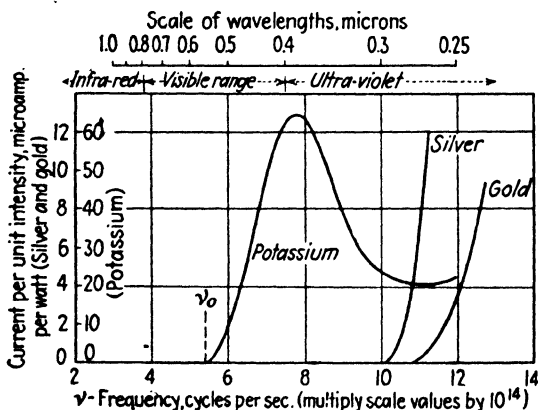


FIG. 6-42.—Spectral distribution curves for clean metals. (Data for K by R. Suhrmann and H. Theising, *Z. Physik*, **52**, 457, 1928; for Ag and Au see general reference 6-8, p. 42.)

experiment is performed with polarized light, it is found that the results depend on the angle of incidence and also on the position of the plane of polarization.

¹ The relation between the wave length λ and the frequency ν is $\nu\lambda = c$, where c is the velocity of light.

The curves in Fig. 6-42 show that the kind of metal of the cathode has an important effect on the emission. The alkali metals and alkaline earth metals alone are sensitive in the region of visible light. Slight contaminations, either in the form of thin surface layers of metal atoms or of adsorbed gases, have an important effect on the spectral distribution curve.

The current through a photoelectric tube can be increased by the introduction of an inert gas at a low pressure. The characteristics of tubes prepared in this manner will be discussed in Sec. 6-13.

b. Theory of the Photoelectric Emission of Electrons.—A study of the velocities of the emitted electrons is particularly helpful toward gaining an understanding of the process by which the electrons are released. The *maximum* velocity of the emitted electrons can be calculated if the negative potential needed to stop all the electrons is known (E_m , Fig. 6-41). Under this condition the initial kinetic energy of the fastest electron is just equal to the energy given up by the electron in the opposing electric field, and therefore

$$\frac{1}{2}mV^2 = E_m q_e \quad (6-78)$$

where V is the initial velocity of the fastest electrons and m is the electronic mass.

Tests using the method described above produced an unexpected result. It was found that the initial velocity of the electrons depended, not on the intensity of the light, but only on the frequency of the light. According to the wave theory of light, one might assume that a more intense light, *i.e.*, light having a higher wave amplitude, would impart a greater energy to the electron. This difficulty was overcome by the introduction of the quantum theory of radiation. This led to the prediction by Einstein, in 1905, that the energy with which the most energetic electrons are emitted is related only to the frequency of the light by the simple relation

$$\frac{1}{2}mV^2 = h\nu - W \quad (6-79)$$

In this expression, ν is the frequency of the light, W is a constant for a given surface, and h is Planck's constant.¹ This simple linear relation has been amply tested experimentally. It has been found to be accurate over wide ranges of frequency.

¹ $h = 6.62 \times 10^{-34}$ joule-sec.

According to the quantum theory of radiation, light consists of discrete, elementary units of radiation called *photons*. The photons seem to have a dual aspect, in that they have both the characteristics of a particle and the characteristics of a wave. The frequency of the light is associated with the wave character of the photons. The photons of a given frequency all have the same energy, but for different frequencies the energy is given by the relation

$$\text{Energy} = h\nu$$

The intensity of the light depends on the product of the number of photons striking the surface per second by the energy of each photon.

Equation (6-79) is therefore interpreted to mean that a photon having an energy $h\nu$ impinges on the surface and as a consequence of its absorption imparts all this energy to one of the free electrons in the surface layers of atoms. The electron loses part of this energy W in emerging from the surface, and the remainder is accounted for by its initial kinetic energy $\frac{1}{2}mV^2$. The energy W is assumed to be required to overcome the same surface attractive forces that played so important a role in thermionic emission. Therefore, the energy W can be identified as the net work function of the surface. By measuring the retarding potential E_m , and thus the maximum initial energy, for a known light frequency, it is possible to determine the work function W . Values measured in this way have been found to agree with the values obtained by thermionic tests for clean metals. The work functions of many substances can be measured by the photoelectric method when it would be impossible to raise the temperature of the material enough to use the thermionic method.

Equation (6-79) also gives a simple explanation of the threshold frequency ν_0 . The frequency ν_0 is obviously that which makes the energy of the photon just equal to the work function, thus:

$$h\nu_0 = W \quad (6-80)$$

Therefore, a low value of work function is accompanied by a low frequency at which the photoelectric effect starts. Most metals have values of ν_0 in the ultraviolet region, but the alkali metals have values of ν_0 that fall in the visible part of the spectrum. The work functions obtained by photoelectric tests for

various clean surfaces and the corresponding threshold frequencies and wave lengths are given in Table 6-10.

TABLE 6-10.—TYPICAL VALUES* OF WORK FUNCTIONS MEASURED BY THE PHOTOELECTRIC METHOD

Surface	Sym- bol	Work function, volts	ν_0 , cycles/sec.	λ_0 , microns†	Color region of ν_0
Tungsten.....	W	4.58	11.2×10^{14}	0.268	Ultraviolet
Cadmium.....	Cd	4.07	9.95×10^{14}	0.301	Ultraviolet
Thorium.....	Th	3.38	8.25×10^{14}	0.364	Ultraviolet
Lithium.....	Li	2.2	5.37×10^{14}	0.558	Yellow
Sodium.....	Na	1.9	4.64×10^{14}	0.647	Orange
Potassium.....	K	1.8	4.4×10^{14}	0.682	Red
Rubidium.....	Rb	1.8	4.4×10^{14}	0.682	Red

* For individual observations, see general reference 6-8, p. 75, and general reference 6-3, Table IV.

† 1 micron = 10^{-6} m. = 10^{-4} cm.

The alkali metals are extremely difficult to prepare in a gas-free condition, and the values of their work function reported by different observers differ by as much as 0.4 volt.

Thus far, the possible effect of the temperature of the emitting surface has been ignored. In the discussion of the distribution in energy of the free electrons according to the Fermi-Dirac laws (Sec. 6-2), we noted that an increase in temperature caused the energies of some of the electrons to be raised above the characteristic energy W_i . Thus it would be expected that a lower threshold frequency ν_0 would be observed as the temperature was increased, for some of the electrons have thermal energies sufficient partly to overcome the work function. However, we saw in Fig. 6-6 that the change in the distribution curve was relatively slight for a large change in the temperature. The detailed theory worked out by Fowler¹ showed that the simpler theory already discussed was strictly true only at absolute zero. His theory showed that at higher temperatures the spectral distribution curve should approach the axis of abscissas tangentially so that the threshold frequency would not be a definite quantity. However, Fowler devised a method of analysis of the experimental results that gave the true threshold frequency, *i.e.*, that

¹ *Phys. Rev.*, **38**, 45, 1931. General reference 6-8, pp. 241ff.

which would be obtained at absolute zero. At room temperature the correction is small, but at about 1000°K . the changes in the apparent threshold are of the order of 5 per cent so that ample experimental verification of the theory has been possible.

An important question from a practical viewpoint concerns the relation between the yield of photoelectrons and the intensity of the incident light. We have seen that the frequency must be above the threshold value ν_0 for the particular surface before an appreciable number of electrons are released. All the photons of light having a higher frequency than ν_0 have sufficient energy to cause the ejection of an electron. There are, however, several reasons why we should not expect *every* incident photon to release an electron. In the first place, only a part of the incident radiation is absorbed. Furthermore, many of the electrons excited by the absorbed photons will have inwardly directed velocities and will take no part in the emission. Another reason lies in the fact that some of the photons are able to penetrate several atomic layers before they transfer their energy to an electron. An electron that receives its energy at an appreciable depth will not be so likely to escape, for it may experience collisions causing energy losses or inward deviations in its path.

The data in Fig. 6-42 make it possible to calculate the number of electrons actually emitted compared with the number of incident photons. Let us take potassium as an example at a frequency of $7.75 \times 10^{14} \text{ sec.}^{-1}$, where its emission is maximum, *i.e.*, $64.5 \times 10^{-6} \text{ amp./watt}$. A current of $64.5 \times 10^{-6} \text{ amp.}$ corresponds to 4.03×10^{14} electrons per second, and 1 watt corresponds to $1/h\nu$ photons per second or 1.95×10^{18} photons per second. Therefore, each photon produces, on the average, 2.06×10^{-4} electron; or, in other words, it requires 4,850 incident photons to release 1 electron. The ratio of the number of electrons to the number of incident photons is called the *quantum yield*. Maximum values of quantum yield as high as 0.0715 have been obtained, *i.e.*, 1 electron for every 14 photons.

As the frequency is increased, the energy of a single photon increases, so that the number of photons in a beam of unit intensity decreases. On this basis, it might be expected that the electron yield would decrease as the frequency is increased. This is actually observed for potassium (Fig. 6-42) and the other alkali metals over certain ranges of frequency.

A complete theory of the quantum yield would, however, need to take all the complex processes of photon absorption and electron excitation into account, and attempts in this direction have met with only partial success.

6-13. Photoelectric Tubes and Their Application.—Photoelectric cathodes composed of a single clean metal are utilized for special purposes, but the majority of phototubes have alkali-metal-covered cathodes whose surfaces are specially prepared with the object of increasing their sensitivity or of extending their color range. These special methods of preparation can be classified as follows: (1) exposure to a glow discharge in hydrogen, (2) exposure to gases containing sulphur, or sulphur and oxygen, (3) the formation of thin films of the sensitive metal on a base metal, and (4) the formation of composite cathodes consisting of a thin film of the alkali metal on a layer of its oxide which, in turn, is formed on an oxide of a base metal. In methods 1 and 2 the gas can be pumped out after the surface has been treated and the alkali-metal surface so formed has a sensitivity many times (order of 5 to 25 times) that of the pure metal. Moreover, the threshold frequency is reduced, a fact indicating a reduction in the work function of the surface. When method 3 is used, *i.e.*, when a thin film of an alkali metal is deposited on a supporting metal such as silver, the work function decreases as the film is made thicker but then increases again. The minimum work function is obtained with films of the order of one or two atoms deep. The photoelectric emission passes through a maximum when the film is about five atoms deep. The behavior of the work function is very similar to the corresponding variation of the work function of thorium on tungsten, measured thermionically, though the maximum emissions occur for quite different film thicknesses in the two cases. The most common example of a composite cathode prepared by method 4 consists of a layer of cesium deposited on oxidized silver. This cathode is indicated by the symbol Ag-CsO-Cs, although the chemical form of the oxides in the intermediate layer is not known with certainty.

The preparation of the sensitive surfaces described in the foregoing paragraph is an art rather than a science,¹ and different tubes made by the same process differ greatly as to sensitivity

¹ For details of techniques, see general reference 6-8, Chap. XII, general reference 6-6, p. 185, and general reference 6-11, Chaps. IV, V.

and threshold. Therefore, spectral distribution curves, such as those shown in Fig. 6-43 for various sensitive surfaces, are subject to wide variations. It should be noted that the curves in Fig. 6-43 are plotted against wave length rather than frequency as in Fig. 6-42. Thus, the threshold frequency corresponds to the longest wave length to which the surface will respond.

The form of spectral distribution curve that is most desirable in a given application depends on the function of the photoelectric tube. If the tube is to be used in the measurement of illumination, its spectral curve should as nearly as possible match the spectral response curve of the human eye. For maximum

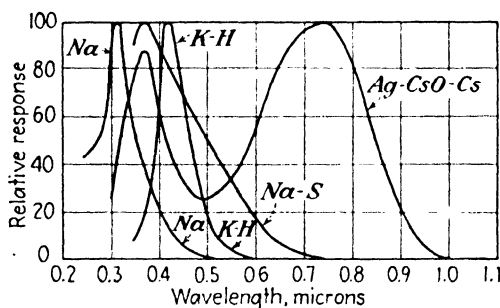


FIG. 6-43.—Spectral distribution curves. Na: pure sodium; Na-S: sodium sensitized with sulphur; K-H: potassium sensitized in a hydrogen glow discharge; Ag-CsO-Cs: cesium deposited on oxidized silver. (General reference 6-10.)

response to light from a given source, say a tungsten lamp, the phototube should respond to the range of wave lengths received from the lamp.

The relative response of the average human eye to light of different wave lengths is shown in Fig. 6-44 and is labeled "visual sensitivity curve." A comparison of the visual sensitivity curve with the spectral curves for various photoelectric surfaces shows that a single tube is unlikely to give a response closely similar to that of the eye. An Na-S or a Cs-Mg surface together with a suitable light filter gives a satisfactory response.

Figure 6-44 also shows the spectral energy distribution curve for a tungsten lamp operated at a color temperature¹ of 2870°K. This curve clearly illustrates why a phototube must be sensitive

¹ A tungsten lamp has a color temperature of 2870°K. when it radiates light within the visible range which has the same spectral distribution as that radiated by a black body at 2870°K.

in the red and infrared regions if it is to have a high sensitivity to light from a tungsten lamp. The Ag-CsO-Cs surfaces are sensitive over a considerable portion of the long-wave-length region and consequently have good sensitivity to this light.

The general shape of the current-voltage curve for a vacuum phototube and the curves for a tube containing one of the inert

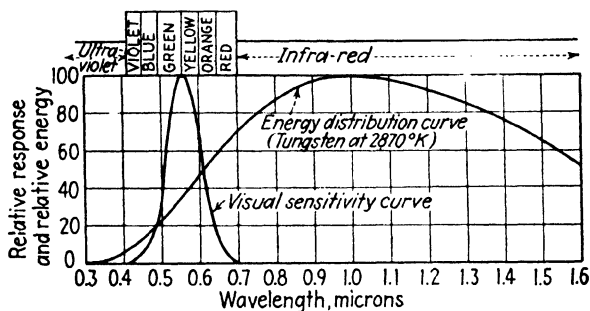


FIG. 6-44.—Visual sensitivity curve and energy distribution curve for the radiation from a tungsten lamp. (General reference 6-11, pp. 26, 29.)

gases under three different gas pressures are shown in Fig. 6-45. The curve for the vacuum photoelectric tube is typical of that in which the anode consists of a small wire. At low voltages, some of the emitted electrons do not strike the anode wire but, instead, travel around the wire in paths resembling hyperbolas. As the voltage is increased, the paths change to spirals which

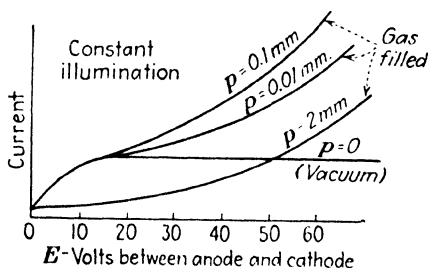


FIG. 6-45.—Current-voltage curves for gas-filled and vacuum phototubes.

converge on the wire, so that finally all the emitted electrons are collected. The current densities are so small that the space charge, which played such an important part in the case of thermionic currents, has little effect.

The increase in the current through the gas-filled tubes shown in Fig. 6-45 is caused by the ionization of the gas which results

from collisions of electrons with the gas atoms. As shown in the figure, the pressure of the gas must be adjusted to rather a critical value, for too high or too low a pressure will result in a reduction in the current. The theory fundamental to the amplification of the photoelectric current by gas ionization will be considered in detail in Sec. 8-5. This method of increasing the sensitivity is used in practical gas-filled photoelectric tubes, and results in a magnification of sensitivity of 5 to 10 times that obtainable with the vacuum phototube. A greater increase in sensitivity cannot be used because a glow discharge starts in the tube when the voltage is increased beyond the safe value. This glow discharge is detrimental to the sensitive surface; and

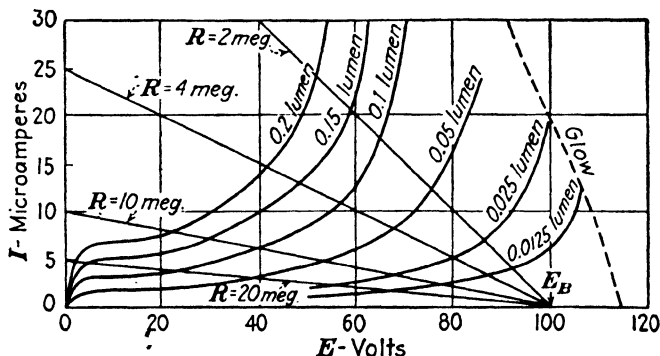


FIG. 6-46.—Voltage-current curves for an argon-filled phototube.

furthermore, the current in the glowing condition is no longer controlled by the light (see Sec. 8-5). The glow voltage depends somewhat on the intensity of the light, as shown in Fig. 6-46. The straight lines in the diagram should be disregarded for the present. As indicated by the dotted line at the right, the glow voltage decreases as the illumination is increased.

The sensitivity rating of a phototube is defined as the *ratio of its current response to the light flux entering the tube when it is illuminated with light from a tungsten filament at a color temperature of 2870°K.* The sensitivity is usually stated in microamperes per lumen.¹ Obviously, this sensitivity rating gives

¹ The units used for measuring the light are those used in illuminating engineering and depend on the visual response of the human eye. The quantity of light, or the *luminous flux*, is measured in *lumens*; the *luminous intensity* of a source is measured in *candles*. The relation between these quantities can be illustrated by considering a hypothetical point source

information relating only to the response to the standard light source, and the response to other light sources can be predicted only if the complete spectral distribution curves of both the phototube and light source are known.

A wide variety of applications has been found for the phototube. It is the essential element in devices that perform such diverse functions as (1) reproducing sound from the sound track on sound-motion-picture film, (2) transmitting pictures over wires or by radio, (3) counting or sorting objects on a production line, (4) turning street lights or illuminated signs on or off at predetermined daylight intensities, and (5) measuring the light output of incandescent lamps. In many devices the phototube serves merely to distinguish between the presence and absence of a beam of light. In others the response must be proportional to the light intensity on the phototube; *i.e.*, the phototube is utilized in the measurement of light intensity. For details regarding the various circuits and systems employed in these applications, as well as for a discussion of photoconductive and photovoltaic cells, the reader is referred to the current literature and the references below.¹

having a luminous intensity of C candles which radiates uniformly in all directions. Then, the luminous flux F in lumens falling on a spherical surface of area S centered on the source is given by

$$F = \frac{CS}{r^2} \quad (6-81)$$

where r is the distance from the source to the surface S . A practical source does not radiate equally in all directions, but in this case Eq. (6-81) can be applied for a single direction away from the source. Then S must be a small portion of the sphere of radius r , and C is the luminous intensity in the direction r .

The *illumination* of a surface is defined as the *density of luminous flux on the surface*; *i.e.*, it is given by the ratio F/S . Representing the illumination by I , we see that the luminous flux is given by

$$F = IS \quad (6-82)$$

In this equation, F is in lumens, and the unit for I is the *foot-candle* when S is measured in square feet. If S is measured in square meters, then the unit of I is the *meter-candle*; or if S is expressed in square centimeters, then I is measured in *photo*.

¹ General references 6-10 and 6-11. HENNY, K., "Electron Tubes in Industry," 2d ed., Chaps. V and VI, McGraw-Hill Book Company, Inc., New York, 1937. REICH, *op. cit.*, Chap. XIII; "Electrical Engineers' Handbook,"

A factor of considerable importance, especially when the phototube is used in the measurement of illumination, is the change in the sensitivity of the tube with time. The slow changes that are observed are particularly noticeable with the various sensitized surfaces, but are nearly absent from the plain metal surfaces. The changes are greater in gas-filled than in vacuum phototubes. The cause of the effect is presumed to be changes in the condition of the sensitive surface, resulting from temperature variations or positive-ion bombardment.

In most photoelectric devices the small currents must be amplified in order to operate a milliammeter or a control relay. The electron multiplier has considerable promise of usefulness in

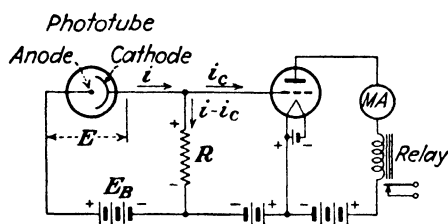


FIG. 6-47.—Phototube amplifier circuit.

this connection (Sec. 6-10). The fundamental circuit diagram of the more commonly employed thermionic amplifiers is shown in Fig. 6-47. When light falls on the phototube, the resulting current through the grid resistor R gives rise to a voltage in the grid circuit of the triode. When the connections are as shown in Fig. 6-47, this voltage raises the grid potential and causes more plate current to flow. When the grid current i_c is negligible, we can write the voltage equation around the loop including E_B , the phototube, and the grid resistance, as follows:

$$E = E_B - iR \quad (6-83)$$

This equation is the same as Eq. (6-52) for the voltage relations in the plate circuit of a grid-controlled thermionic tube. Therefore, the graphical analysis employed in Sec. 6-11 can also be used here. Thus, the straight lines in Fig. 6-46 show possible phototube-current and -voltage values corresponding to the assumed battery voltage and the various values of R . The intersection

of a particular resistance line with a tube voltage-current curve gives the tube current and voltage at that particular value of light flux. For example, when $E_b = 100$ volts, $R = 2$ megohms, and the light flux is 0.1 lumen, the current will be $17.4 \mu a$ and the voltage across the tube will be 65 volts. Consequently, the voltage drop across R is 35 volts. As the light flux increases, the current through the tube increases in a regular although not in a linear fashion.

The effect of a change in the value of the grid resistance R is easily studied by inspecting the various resistance lines in Fig. 6-46. It is seen that fairly high values of resistance are needed, especially when the light is weak. In extreme cases, as in the photometry of the stars, the insulation resistance of the circuit or the grid current of the thermionic tube may be the limiting factor that determines the maximum permissible value of grid resistance. When high resistances are employed, the current response approaches a limiting value for large light intensities. A study of Fig. 6-46 will show that this would occur for the grid resistance of 20 megohms at values of light flux over about 0.15 lumen.

The analysis of the behavior of a vacuum phototube can be carried out by the same method as that illustrated in Fig. 6-46 for the gas-filled tube. In this case, however, the higher voltage portions of the current-voltage curves of the tube are equally spaced horizontal lines. As a result, the response is linearly related to the quantity of light as long as the voltage across the tube remains above the saturating value.

An interesting observation on gas-filled phototubes is that an appreciable time elapses after a sudden change in illumination before the current through the tube reaches a steady value. It has been established that there is a negligible time lag between the incidence of the light and the ejection of the photoelectrons, so that the observed lag must be attributed to the time required for the gaseous ionization to build up to a steady value. This is partly determined by the time required for the relatively slow gaseous ions (see Sec. 7-1) to cross the tube. Another aspect of the time lag is encountered when a varying beam of light falls on the phototube. This situation exists in the sound-motion-picture projector where the variations in the light passing through the sound track are detected by a phototube and after

amplification are converted into sound in a loud-speaker. The current through the phototube does not correspond exactly with the light variation when the frequency of the variation is above a few cycles per second. The decrease in the amplitude of the response from zero frequency to 5,000 cycles/sec. is of the order of 20 per cent. This decrease is small enough so as not to bar the gas-filled tube from the sound-motion-picture industry; in fact, the gas-filled tubes are most often employed.

6-14. Secondary Electron Emission Caused by Positive-ion Impact.—When positive ions bombard a negatively charged electrode—as at the cathode in a glow discharge—it is important to know whether or not electrons are released by these impacts. We have seen (Sec. 6-10) how electron bombardment releases secondary electrons; we now consider the analogous process of secondary electron release caused by positive-ion bombardment. Under conditions such as those in the glow discharge the phenomenon is often complicated by the presence of electron emission caused by the impacts of metastable atoms or by the photoelectric effect. In the most reliable experiments, therefore, beams of ions are shot into a high-vacuum tube, and the electrons released as the beam strikes a metal target are collected and measured. A correction must be made for the current carried by the positive ions that are reflected from the target.

The average yield of secondary electrons per impinging positive ion is expressed by the symbol γ . Thus γ is the ratio of the number of free electrons released to the number of impinging positive ions. It is also the probability that an ion will release an electron. Relatively few quantitative investigations of the factors affecting γ are available; however, it has been shown that γ depends on (1) the energy and angle of incidence of the bombarding ions, (2) the kind of metal of the target and its surface condition, and (3) the nature of the ions.

The types of variation of the secondary-electron yield with the energy of the ions are shown in Fig. 6-48. By comparing these curves with those in Fig. 6-27, it will be noted that, in general, the yields are less than those obtained owing to the impact of electrons of equal energy. The effect of the condition of the metal surface is shown by the curves for Cs^+ ions bombarding hot Ni and cold tarnished Ni. A similar effect is evident in the two curves for He^+ ions, as obtained by Oliphant. These

results indicate that degassing the target reduces the yield of secondary electrons.

It has been suggested that in the case of low-velocity ions the potential energy of the ion may account for the release of a free electron. For example, if its potential energy were great enough, the ion could pull two electrons out of the metal, one of which would go to neutralize the ion and the second of which would become free. If in this process other gains or losses are neglected, then the potential energy¹ of the ion would have to exceed twice the net work function of the metal for emission to take place. This restriction would not apply to the high-velocity ions, for

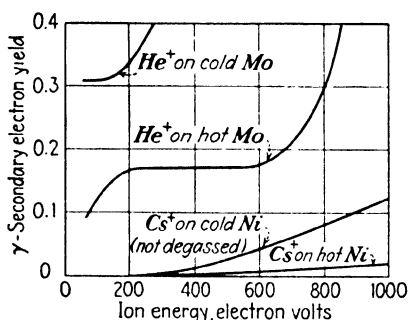


FIG. 6-48.—Secondary-electron emission caused by positive ion bombardment. [Data for He⁺ by M. L. E. Oliphant, *Proc. Roy. Soc. (London)*, **127**, 379, 1930; for Cs⁺ by J. Kock, *Z. Physik*, **100**, 700, 1936, and P. B. Moon, *Camb. Phil. Soc. Proc.*, **27**, 573, 1931.]

these can transfer a small proportion (of the order of the ratio of the electronic mass to the ionic mass) of their kinetic energy to an electron in the metal. This electron will subsequently be emitted if its path is turned outward by elastic collisions and if it has enough energy to overcome the surface forces.

Problems

1. (a) Calculate the inner work function of metallic thorium on the assumption of one free electron per atom. The atomic weight of thorium is 0.232 kg., and its density is 11,200 kg./cu. m. (b) If the net work function of thorium is 3.38 e.v., what is its outer work function?

2. If point C in Fig. 6-8 is 0.5×10^{-9} m. from the surface, what is the contribution of the image force to the outer work function in the range from C to infinity? Express the result in electron volts.

3. Show that the arithmetical average speed of the electrons in a metal at 0°K. on the basis of the Fermi distribution law is three-fourths of the maximum velocity, $V_{\max} = \sqrt{2W_i/m}$.

¹ See Sec. 8-1 for a discussion of ion energies.

4. (a) Show that the total number of electrons which cross a plane of unit area within a metal per second, for the Fermi distribution at $T = 0^\circ\text{K.}$, is given by

$$\eta = \frac{2\pi m}{h^3} W_i^2$$

where W_i is the inner work function of the metal. (b) Calculate the value of η of part *a* for tungsten on the assumption of one free electron per atom (see Table 6-1).

5. (a) A 5-kw. high-vacuum triode has a tungsten filament that consists of two hairpin loops each 9.5 cm. long. If the effective total length of the filament wire is 38 cm. and its diameter is 1 mm., what is the saturation emission from this filament at 2500°K. ? (b) If the cathode efficiency of this filament is 5 ma./watt, what heating power is required for the filament in part *a*?

6. (a) What is the ratio of the emission from a tungsten filament at 2400°K. to that at 2500°K. ? (b) What conclusion can be drawn from the result of part *a* as to the permissible variations of tungsten-filament temperature in operation?

7. The constants in the emission formula for a given thoriated-tungsten filament are $A = 1.5 \text{ amp./cm.}^2(\text{deg.})^2$ and $\phi_0 = 2.63 \text{ e.v.}$ What is the saturation emission from this filament at 1900°K. ?

8. Compare the cost of operating 50,000 telephone repeater tubes for 9,000 hr. when they have filaments of (a) tungsten, (b) thoriated-tungsten, and (c) oxide-coated nickel. Assume that an emission current of 100 ma. is required and that the cathode efficiencies are as follows: 4 ma./watt for tungsten, 30 ma./watt for thoriated-tungsten, and 50 ma./watt for oxide-coated nickel. The cost of power is 5 cents per kilowatt-hour.

9. (a) By what amount is the work function of tungsten reduced by an external field intensity of 200 million volts per meter? (b) Would the answer to part *a* be different for a different metal? (c) With reference to part *a*, at what distance from the surface does an electron experience no force, it being assumed that the image-force law holds in this region?

10. In a certain experimental high-vacuum diode the electron emission from the tungsten filament is saturated at a voltage of 100 volts on the anode and the emission current is 0.100 amp. If the electric intensity at the cathode corresponding to an anode voltage of 100 volts is 10^6 volts/m. and the temperature of the filament is 2450°K. , what is the emission current when the anode voltage is 3,600 volts?

11. A space-charge-limited electron current flows between a plane cathode and a parallel-plane anode. Assuming that the electrons are given off by the cathode with zero velocity, find the mathematical form of (a) the variation of the electric intensity at any point between the electrodes. Similarly, find the variations of (b) space-charge density and (c) the electron velocity in the region between the electrodes. Sketch curves showing the nature of these variations, and include a curve showing the potential distribution between the electrodes.

12. (a) What anode voltage would be required to cause a space-charge-limited electron current of 0.1 amp./sq. cm. between a plane hot cathode

and a cold anode separated by 2 mm.? (b) What voltage would be required if the separation were 10 mm.?

HINT: Assume the electrons are emitted from the cathode with zero velocity.

13. What space-charge-limited current (zero initial velocity being assumed) will be produced by an anode voltage of 150 volts in a diode consisting of a wire filament down the center of a cylindrical anode, if the filament is 3 cm. long and has a diameter of 0.2 mm. and the inner radius of the anode is 10 mm.?

14. (a) The relation between current and voltage for a given diode is $I = 10^{-4}E^{\frac{3}{2}}$. Plot the current-voltage curve up to a voltage of 150 volts. (b) Using the curve from part a, derive the wave form of the current through the diode when a sinusoidal alternating e.m.f. of 180 peak volts is impressed on the diode through a resistance of 800 ohms.

15. (a) In a certain triode the plate current falls to zero when $E_c = -18$ volt and $E_b = 250$ volts. Estimate the value of μ of the tube. (b) Another test shows that the plate current of this tube is 0.005 amp. when $E_c = -10$ volt and $E_b = 200$ volts. Assuming that the three-halves-power law holds and that μ has the value found in part a, find the constants in the empirical relation of plate current to grid and plate voltages.

16. A triode has a filamentary cathode with a concentric helical grid and cylindrical anode. The grid helix has a radius of 4 mm., with 11 turns per centimeter, and the grid-wire diameter is 0.10 mm. If the inner radius of the anode is 7 mm., what is the approximate value of the μ of this triode?

17. In a certain electron multiplier the yield ratio for secondary emission is 7.5. If the multiplier has six stages and there is no loss from the electron beams, what is the amplification of the photoelectric current?

18. The triode whose plate characteristics are shown in Fig. 6-24 is to be operated in a simple amplifier circuit similar to that shown in Fig. 6-31. (a) Draw a circuit diagram for this amplifier. (b) If the plate supply voltage E_B is 300 volts, the grid bias voltage E_c is -10 volt, and the plate current with no signal is to be 5 ma., what is the value of the load resistance? (c) What is the voltage amplification for small signals under the conditions of part b?

19. (a) Calculate the power output of the amplifier analyzed in Fig. 6-32 for a peak input signal voltage of 5 volts. (b) Under these conditions, what is the power expended in heating the plate? (c) What power is expended in heating the plate when the input voltage is zero?

20. Light having a frequency of 5×10^{14} sec.⁻¹ falls on a photoelectric cathode. If a retarding potential of 0.393 volt (corrected for contact potential) reduces the electron current to zero, what is the work function of the cathode surface?

21. What is the photoelectric threshold frequency for a zinc surface whose work function is 3.4 volts?

22. If green light from a mercury arc having a wave length of 0.546 micron shines on a potassium surface, what will be the maximum kinetic energy of the emitted electrons? Express the result in electron volts.

23. Two monochromatic light beams appear to the eye to be equally intense. One has a wave length of 0.555 micron, and the other has a wave

length of 0.65 micron. What relative response would these beams of light produce in a tube containing the Ag-CsO-Cs surface whose spectral curve is shown in Fig. 6-43?

24. (a) What voltages will result across the resistance R in Fig. 6-47 when the quantity of light entering the phototube is 0.05, 0.10, and 0.15 lumen, it being assumed that the characteristics of the phototube are those given in Fig. 6-46, that the voltage E_B is 100 volts, and that R is 10 megohms? (b) What is the illumination in foot-candles needed in each case to produce the fluxes mentioned in part *a* if the cathode of the phototube has an effective area of 0.9 sq. in.?

25. Choose a value of the resistance R in Fig. 6-47 and of the grid bias voltage on the triode so that the relay will close when the light flux on the phototube reaches 0.05 lumen but will open when the light is cut off. Assume that the value of E_B is 70 volts, that the plate supply voltage for the triode is 200 volts, that the resistance of the relay coil is 1,200 ohms, and that the characteristics of the triode are those shown in Fig. 6-24. The relay requires 11 ma. to close and opens when the current falls to 4 ma. The phototube characteristics are those in Fig. 6-46.

General References

- 6-1. REIMANN, A. L.: "Thermionic Emission," John Wiley & Sons, Inc., New York, 1934.
- 6-2. CHAFFEE, E. L.: "Theory of Thermionic Vacuum Tubes," McGraw-Hill Book Company, Inc., New York, 1933.
- 6-3. BECKER, J. A.: "Thermionic Electron Emission," *Rev. Mod. Phys.*, **7**, 95, 1935; *Bell System Tech. J.*, **14**, 413, 1935.
- 6-4. DUSHMAN, SAUL: "Thermionic Emission," *Rev. Mod. Phys.*, **2**, 381-476, 1930.
- 6-5. JONES, T. J.: "Thermionic Emission," Methuen & Co., Ltd., London, 1936.
- 6-6. KOLLER, L. R.: "The Physics of Electron Tubes," 2d ed., McGraw-Hill Book Company, Inc., New York, 1937.
- 6-7. DOW, W. G.: "Fundamentals of Engineering Electronics," John Wiley & Sons, Inc., New York, 1937.
- 6-8. HUGHES, A. L., and L. A. DUBRIDGE: "Photoelectric Phenomena," McGraw-Hill Book Company, Inc., New York, 1932.
- 6-9. LINFORD, L. B.: "Recent Developments in the Study of the External Photoelectric Effect," *Rev. Mod. Phys.*, **5**, 34-61, 1933.
- 6-10. CAMPBELL, N. R., and D. RITCHIE: "Photoelectric Cells," 3d ed., Sir Isaac Pitman & Sons, Ltd., London, 1934.
- 6-11. ZWORYKIN, V. K., and E. D. WILSON: "Photocells and Their Applications," 2d ed., John Wiley & Sons, Inc., New York, 1932.
- 6-12. DE BOER, J. H.: "Electron Emission and Adsorption Phenomena," The Macmillan Company, New York, 1935.
- 6-13. COMPTON, K. T., and I. LANGMUIR: "Electrical Discharges in Gases" *Rev. Mod. Phys.*, **2**, 123, 1930; **3**, 191, 1931.

CHAPTER VII

DEIONIZING PROCESSES¹

7-1. Electric Field—Mobility.—Among the deionizing agents, the electric field is probably the most important, for in any case involving the flow of current through a gas we shall always have an electric field moving the ions toward one electrode or the other. In Chap. I, we saw that the current density in a discharge tube was given by an expression of the form,

$$J = N_1 q_1 \bar{V}_1 + N_2 q_2 \bar{V}_2 + \dots \quad (1-2)$$

In this expression, N_1 and N_2 were the densities of ions of type 1 and 2, q_1 and q_2 were the quantities of charge carried by each ion, and $\bar{V}_1$ and $\bar{V}_2$ were the average velocities with which the ions moved in the direction (or counter to the direction) of the electric field. In any particular problem, then, we see that, although the electric field has some influence on the ion densities, the principal effect of the electric field on the current flowing is felt in the variation it causes in the average drift velocity $\bar{V}_1$ or $\bar{V}_2$. In the case of a high-vacuum discharge, this dependence was quite simple because there were no collisions with neutral gas atoms to complicate matters. When collisions with neutral atoms or molecules occur, these must be taken into account and the problem immediately becomes more involved.

Because the dependence of drift velocity on field strength or potential is so essential to the problem of determining the relationship between discharge current and voltage, it will be worth while to examine briefly the manner of this dependence when the gas pressure is not negligible. Let us say that the drift velocity is directly proportional to the field strength as shown in Eq. (7-1),

$$\bar{V} = K F \quad (7-1)$$

Then we can define the *mobility* K of an ion in an electric field as the ratio of its average drift velocity to the field strength.

¹ General references for Chap. VII will be found on p. 258.

$$K = \frac{V}{F} \quad (7-2)$$

We see that, if V is expressed in meters per second and F is expressed in volts per meter, K will be expressed in terms of meters per second per volt per meter. The mobility serves a useful function in that it enables us to predict the order of magnitude of ionic velocities or discharge currents in a given electric field. Order-of-magnitude calculations are usually sufficient when we are interested only in trends. For particular applications, we shall probably find it best to use empirical data in much the same way as in many engineering problems.

In general, the mobility is dependent upon the field strength and is not really a constant. This may perhaps be a little confusing at first, but there are many instances in engineering where we define "constants" so that they can be used in calculations when, as a matter of fact, these quantities are not actually constant except under certain conditions. A single illustration should suffice. In a magnetic circuit we say that the flux density is proportional to the magnetic intensity.

$$B = \mu H$$

The proportionality constant μ , however, is constant only under certain conditions—in an air-core circuit, for example. In an iron-core circuit, the permeability μ depends upon the magnetic intensity H , as well as on the previous treatment of the core. In much the same way, we shall see that the mobility is constant under some circumstances but in others depends strongly on the electric intensity.

As a first approximation in the calculation of K , assume that we are dealing with a high-pressure discharge in which the field strength is relatively small. Such a condition would be encountered in a d.c. arc at atmospheric pressure. Assume, also, that we are dealing only with singly charged molecular ions. Under these conditions the energy gained by an ion over a free path due to the electric field will be small compared with the average energy of thermal agitation of the molecules, and the motions of the ions will be only slightly different from the motions of the neutral molecules. That is, during each free path the ion gains a small component of velocity ΔV due to the electric field, but the

change in velocity of the ion from V_1 to V_2 due to ΔV is so small that, as far as calculations of velocity are concerned, it can be neglected.

The relationship between V_1 and V_2 is illustrated in Fig. 7-1 for two different directions of motion of an ion. In each case, V_1 represents the velocity at the beginning of a free path, and V_2 represents the velocity of the ion at the end of the free path. The change in velocity due to the electric field F is represented by ΔV . The net effect of the small velocity component ΔV gained on each free path is that the ion progresses slowly in the direction of the field, making thousands or millions of collisions before advancing 1 cm. A projection of its path on a plane parallel to F might then appear as in Fig. 7-2. The motion of the ions is thus very much as if the field were not present. The predominant motion of the ions is random, as in the case of no field whatever; but superimposed on this motion is a slow drift in the direction of the field.

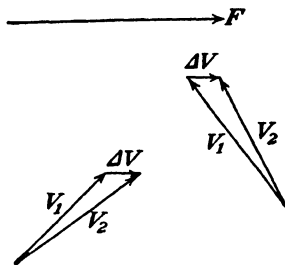


Fig. 7-1.—Change of ion velocity in an electric field.

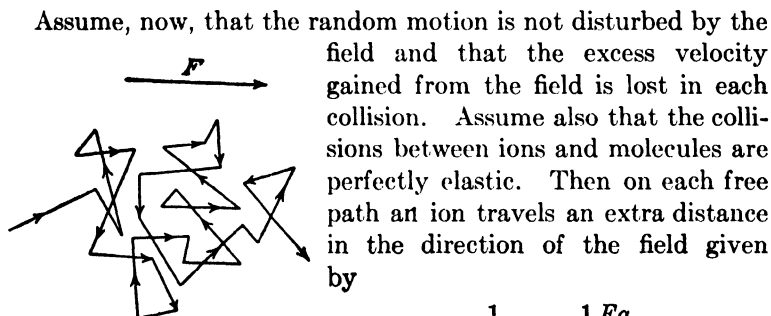


Fig. 7-2.—Progression of an ion in an electric field.

Assume, now, that the random motion is not disturbed by the field and that the excess velocity gained from the field is lost in each collision. Assume also that the collisions between ions and molecules are perfectly elastic. Then on each free path an ion travels an extra distance in the direction of the field given by

$$s = \frac{1}{2}at^2 = \frac{1}{2}\frac{Fq}{m}t^2 \quad (7-3)$$

If as a first approximation all the free paths are taken equal to the mean free path, then the average time between collisions is $t = \lambda/\bar{c}$, $\bar{c}$ being the arithmetical average thermal velocity, and the average distance traveled in the direction of the field is

$$s = \frac{1}{2}\frac{Fq}{m}\frac{\lambda^2}{\bar{c}^2}$$

The average velocity of drift is then

$$\bar{V} = \frac{s}{t} = \frac{1}{2} \frac{Fq}{m} \frac{\lambda}{\bar{c}}$$

Thus,

$$K = \frac{1}{2} \frac{q\lambda}{m\bar{c}}$$

Taking into account the fact that the free paths and the time intervals between collisions are not equal but are distributed at random, the expression given above for the mobility is modified by a factor of 2. The more exact derivation gives

$$K = \frac{q\lambda}{m\bar{c}} \quad (7-4)$$

The reason for this follows directly from a consideration of the distribution of free paths. If in Eq. (7-3) we set $t = x/\bar{c}$ where x is the actual length of a free path, we obtain

$$s = \frac{1}{2} \frac{Fq}{m} \frac{x^2}{\bar{c}^2}$$

To obtain the average distance moved in the direction of the electric field, we must now average over a large number of free paths. The number of free paths with lengths between x and $x + dx$ is given by Eq. (4-65), Chap. IV,

$$|dn| = \frac{n_0}{\lambda} e^{-\frac{x}{\lambda}} dx \quad (4-65)$$

where n_0 is the total number of free paths and λ is the mean free path. Thus the average distance moved in the direction of the field becomes

$$s = \frac{1}{2} \frac{Fq}{m\bar{c}^2} \frac{\int_0^\infty x^2 \frac{n_0}{\lambda} e^{-\frac{x}{\lambda}} dx}{n_0}$$

or, integrating,

$$s = \frac{Fq\lambda^2}{m\bar{c}^2}$$

The average time for traveling one free path is still $\lambda/\bar{c}$; thus,

$$\bar{V} = \frac{s}{t} = \frac{Fq\lambda}{m\bar{c}}$$

and

$$K = \frac{q\lambda}{m\bar{c}} \quad (7-4)$$

If we take into account the fact that the masses of the ions and the molecules are not the same and that the slight amount of energy gained by the ions in the electric field is not all lost to the neutral molecules in the successive collisions, a more general

expression for the mobility can be obtained. This equation was derived by Thomson on the basis of the kinetic-theory relationship between mobility and the coefficient of diffusion. It is a perfectly general relationship valid under all conditions of field strength regardless of the size of the ions and has the following form:

$$K = \frac{0.921q\lambda_0}{mC_m} \sqrt{1 + \frac{m\bar{W}_M}{M\bar{W}_m}} \quad (7-5)$$

In this equation, $\bar{W}_m$ and $\bar{W}_M$ are, respectively, the average kinetic energies of the ions and neutral gas molecules, m and C_m refer to the mass and r.m.s. velocity of the ions, and M refers to the mass of the neutral molecules. λ_0 is proportional to the mean free path of the ions and is defined by

$$\lambda_0 = \lambda \sqrt{1 + \frac{\bar{W}_M m}{\bar{W}_m M}} \quad (7-6)$$

where λ is the kinetic-theory mean free path of the ions. From this we see that when we are considering the motions of electrons, $\lambda_0 = \lambda_e$; but when we are considering the motions of molecular ions, $\lambda_0 = \sqrt{2} \lambda_{\text{molecule}}$, approximately.

When the field strength is very small and the average energy of the ions is approximately equal to the average energy of the neutral molecules, Eq. (7-5) reduces to Eq. (7-4), derived above, except for a small numerical factor. Under these conditions the mobility is independent of the field strength. When the field strength is appreciable, the average kinetic energy of the ions will be larger than the average kinetic energy of the neutral molecules, for the ions will be gaining energy from the electric field which they do not entirely lose in collisions with neutral molecules. The mobility under these conditions thus depends on the field strength, and in order to determine this dependence we shall have to find values of $\bar{W}_m$ and C_m to use in Eq. (7-5).¹

The fraction of the energy of the ion that is lost in a collision with a molecule is given by²

$$f = 2.66 \frac{mM}{(m+M)^2} \left(1 - \frac{\bar{W}_M}{\bar{W}_m}\right) \quad (7-7)$$

¹ The following derivation is based on the works of Hertz (*Z. Physik*, **32**, 298, 1925; *Physica*, **2**, 15, 1922) and Compton (*Phys. Rev.*, **22**, 333, 1923).

² CRAVATH, A. M., *Phys. Rev.*, **36**, 248, 1930.

This is a very small number in the case of electrons moving through a gas; for then $m \ll M$ and $\bar{W}_m \gg \bar{W}_M$, and f is approximately equal to $2.66m/M$. Thus, the energy of an ion starting from rest in the electric field gradually increases between successive collisions, but it does not continue to increase indefinitely. Eventually a condition is reached where, on the average, just as much energy is lost in a collision as is gained in the electric field during the next free path. When this condition is reached, the ion has an average energy given by

$$W_m^* = \frac{qF}{\zeta f} \quad (7-8)$$

where W_m^* refers to the *terminal energy* of the ion and ζ is the average number of collisions made by the ion while advancing unit distance.¹

To evaluate ζ , we note that the number of collisions made by an ion in unit time is given by $\bar{c}_m/\lambda$. Thus, the number of collisions made by the ion in advancing unit distance in the direction of the field is

$$\begin{aligned} \zeta &= \frac{\bar{c}_m/\lambda}{\bar{V}} \\ &= \frac{2W_m^*}{q\lambda\lambda_0 F \sqrt{1 + \frac{m\bar{W}_M}{MW_m^*}}} \end{aligned}$$

¹ This equation can be deduced as follows: The energy gained by the ions in moving a distance dx in the direction of the field is

$$d\bar{W}_m = qF dx$$

The energy lost in moving a distance dx owing to collisions with neutral atoms is $f\bar{W}_m\zeta dx$. Thus, the net rate of gain of energy of the ions per meter is given by

$$\frac{d\bar{W}_m}{dx} = qF - f\zeta\bar{W}_m$$

The terminal energy can be obtained by equating this to zero; for then the rate of gain of energy is equal to the rate of loss of energy, and the net gain of energy is zero. We see then that since

$$qF - f\zeta W_m^* = 0$$

we have

$$W_m^* = \frac{qF}{\zeta f}$$

since $\bar{V}$ is given by Eqs. (7-1) and (7-5) and since $\bar{W}_m$ is now given by W_m^* . Substituting in this equation the value of λ given by Eq. (7-6), we have

$$\zeta = \frac{2W_m^*}{q\lambda_0^2 F} \quad (7-9)$$

We see then that since $1/\lambda_0$ is proportional to the number of collisions made per unit distance along the path of the ion, ζ , the number of collisions made in advancing 1 m. in the direction of the field is proportional to the square of the number of collisions made in going 1 m. along the *actual* path of the ion. The values of ζ from Eq. (7-9) and f from Eq. (7-7) may now be substituted in Eq. (7-8) and the resulting equation solved for W_m^* . The result of this calculation is given in Eq. (7-10). We have now established the dependence of W_m^* on the field strength F .

$$W_m^* = \frac{\bar{W}_M}{2} + \left[\frac{\bar{W}_M^2}{4} + \frac{q^2 F^2 \lambda_0^2 (m + M)^2}{5.32 m M} \right]^{\frac{1}{2}} \quad (7-10)$$

This equation, of course, reduces to $W_m^* = \bar{W}_M$ when the field strength is very small. When F is very large, this equation shows us that W_m^* is directly proportional to F .

We now have sufficient information to evaluate the mobility K . Remembering that $\frac{1}{2} m C_m^2 = W_m^* = \bar{W}_m$, we can now substitute for C_m and $\bar{W}_m$ in Eq. (7-5) and obtain the following expression for the mobility:

$$K = \frac{0.921 q \lambda_0}{\sqrt{2m}} \left\{ \frac{\bar{W}_M}{2} + \left[\frac{\bar{W}_M^2}{4} + \frac{q^2 F^2 \lambda_0^2 (m + M)^2}{5.32 m M} \right]^{\frac{1}{2}} \right\} \times \left\{ 1 + \frac{m}{M} \frac{1}{\left\{ \frac{1}{2} + \left[\frac{1}{4} + \frac{q^2 F^2 \lambda_0^2 (m + M)^2}{5.32 m M \bar{W}_M^2} \right]^{\frac{1}{2}} \right\}} \right\} \quad (7-11)$$

The result of expanding the expression for the mobility to take account of the field strength is thus the equation above. This is a perfectly general equation applying under all conditions of field strength, regardless of the size of the ions.¹

Under special conditions, Eq. (7-11) can be reduced to much simpler forms. When the field strength is very low, the terms involving F can be neglected and the equation reduces to

¹ $\bar{W}_M$ is, of course, dependent only on the gas temperature (see Sec. 4-6).

$$K_{\text{low field}} = \frac{0.921q\lambda_0}{\sqrt{2m}} \frac{1}{(\bar{W}_M)^{\frac{1}{2}}} \left(1 + \frac{m}{M}\right)^{\frac{1}{2}} \quad (7-12)$$

or since, with a very low field strength, $\bar{W}_m = \bar{W}_M$, we have

$$K_{\text{low field}} = \frac{0.921q\lambda_0}{mC_m} \left(1 + \frac{m}{M}\right)^{\frac{1}{2}} \quad (7-13)$$

Except for a numerical factor slightly different in the cases of electrons and molecular ions, this is the same as Eq. (7-4) which was derived above on somewhat simplified assumptions. In the case of a very high field strength, the terms involving $\bar{W}_M$ in the first bracketed factor can be neglected, and the last product factor in Eq. (7-11) can be set equal to 1. Thus, the equation reduces to a much simpler form given by

$$K_{\text{high field}} = 0.99 \left(\frac{M}{m}\right)^{\frac{1}{2}} \left[\frac{q\lambda_0}{(m+M)F} \right]^{\frac{1}{2}} \quad (7-14)$$

In using these equations, we must bear in mind that λ_0 is not the actual mean free path of the ions but is defined by Eq. (7-6). We can see from these equations that, although in the case of a very small electric field the mobility is independent of the field strength, in the case of a strong electric field the mobility is inversely proportional to the square root of the field strength. The reason for this, of course, depends upon the fact, already pointed out above, that in a strong electric field the ions are gaining a large amount of energy from the electric field on each free path and consequently the ions will not move about in the same way that they do in a field-free region—their average velocities will be much higher than the average velocities of neutral molecules.

This point is further emphasized by the criteria determining whether we can use the approximate equations (7-13) and (7-14) or whether we must use the more exact equation (7-11). An examination of the method of derivation of the high- and low-field forms of the mobility equation shows that we can define a low field as one in which

$$\bar{W}_M \gg qF\lambda_0 \frac{(m+M)}{\sqrt{mM}} \quad (7-15)$$

and that we can define a high field as one in which

$$\bar{W}_M \ll qF\lambda_0 \frac{(m+M)}{\sqrt{mM}} \quad (7-16)$$

Since $qF\lambda_0$ is approximately the energy gained by an ion in moving a distance in the direction of the field equal to the mean free path, these inequalities give a measure of the importance of the energy gained in the electric field compared with the thermal energy in determining the motions of the ions.

Some idea of the exactness of the expression for the mobility derived above can be obtained by comparing the experimental and calculated values of the mobility of electrons in nitrogen and hydrogen. In Fig. 7-3 the calculated value of the mobility is represented by curve 1 and the experimental values by curve 2.

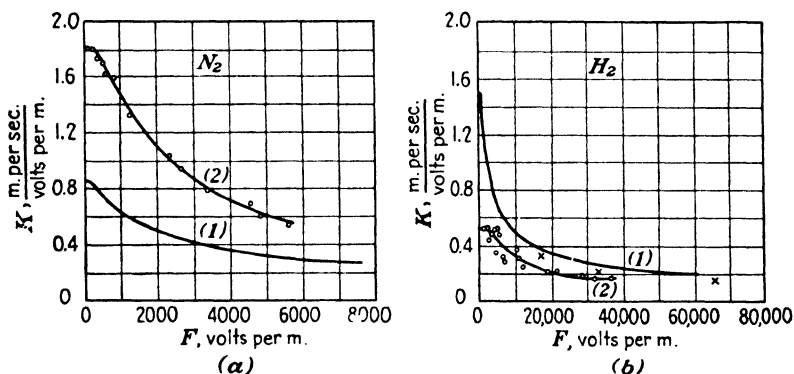


FIG. 7-3.—Mobility of electrons under standard conditions of temperature and pressure. (*K. T. Compton and I. Langmuir, Rev. Mod. Phys.*, **2**, 235, 1930.)

Both curves give the mobility under standard conditions of temperature and pressure. The discrepancies that still exist between the experimental and calculated values can be largely explained as due to the facts that the mean free path is not constant but varies quite widely with the velocity of the ions and that the ions when colliding with neutral molecules are scattered more strongly in the forward direction than calculations on the basis of elastic spheres predict. The theory is adequate, however, to determine the order of magnitude of the mobility under any given condition, and it certainly predicts the dependence on field strength as shown in Fig. 7-3.

From the theory, since K is proportional to the mean free path under conditions of low field strength, Kp should be a constant. This has been experimentally verified for most gases for pressures ranging from 75 atm. down to about 100 mm. Above about 75 atm. the mobility of negative ions decreases, and below 100

mm. it increases rapidly. One would conclude from this that at high pressures many of the ions consist of several molecules closely bound together and acting as a unit, the mass of the ion being thus increased, and that at very low pressures the negative ions exist largely as free electrons, the very small mass producing the large mobility.

There are a number of methods for determining the mobility. Two will be described to show the general type of experiment.

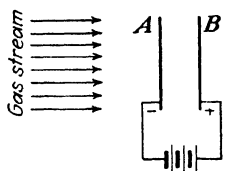


FIG. 7-4.—Zeleny method for comparing mobilities.

One method used by Zeleny consists in having a current of gas moving between two parallel electrodes of wire gauze (Fig. 7-4). The gas is uniformly ionized between the electrodes and moving with a velocity V . If K_+ is the mobility of the positive ions and the field necessary to cause them to move against the gas stream to the cathode is F_+ , the velocity with which they move toward A is $K_+F_+ - V$. Now, by reversing the battery connections the velocity of the negative ions against the gas blast is $K_-F_- - V$. If F_+ and F_- are adjusted so that the velocity of either kind of ion against the gas stream is the same (determined by the field strengths when the currents collected by A are the same), then the ratio of the mobilities of the positive and negative ions can be determined.

$$K_+F_+ - V = K_-F_- - V$$

and

$$\frac{K_+}{K_-} = \frac{F_-}{F_+} \quad (7-17)$$

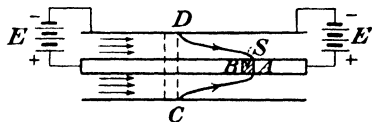


FIG. 7-5.—Rutherford method for determining mobility.

A method of obtaining the absolute value of the mobility is due to Rutherford. A diagram of the apparatus is represented in Fig. 7-5. The method consists in using two concentric cylinders, the inner of which is cut at A so that insulation S may be inserted between the two halves of the cylinder. A beam of electrons is passed through the gas at CD, or the gas in this region is ionized by means of X rays. This ionized stratum of the gas is then caused to move down the tube with a uniform velocity by means of a gas blast. A potential between A and D will then cause a current to flow if the velocity of the gas is high enough to carry ions beyond the insulation. When the velocity

is just large enough so that a current begins to flow, that means that a free ion or electron near the outer cylinder at *D* or *C* will be carried just beyond the insulation before it is drawn into the inner cylinder by the action of the electric field. All other ions formed between *C* and *D* will be drawn into *B* and will produce no current. Under this condition, if the velocity of the gas and the field strength are known, the mobility can be calculated.

7-2. Diffusion.—Diffusion, or the movement of ions solely on the basis of their thermal motions, is one of the two most important means for the removal of ions from a discharge, the other being the electric field. At low pressures and in any discharge with a small electrode separation, the electric field is most important; but in many discharges, especially at high pressure, diffusion accounts for the greatest loss of ionization from the discharge region. For example, in a neon sign most of the deionization is caused by diffusion to the tube walls, and the fact that a high voltage is necessary to maintain the discharge is due to the necessity of continually regenerating the ionization in order that conduction can occur. Another example of the importance of diffusion as a deionizing agent occurs in the case of the extinction of an a.c. arc. As the arc current passes through zero, the voltage also passes through zero and the electric field loses all its effectiveness both as an ionizing and as a deionizing agent. However, diffusion continues to take place, and the ion density continues to decrease because of it. If the diffusion process has been efficient enough in removing ionization, then, when the voltage is reapplied, there will be an insufficient number of ions to carry the current and the discharge will have been interrupted. This process is thus of considerable importance in the design of circuit breakers or high-voltage fuses.

Since the process of diffusion does not depend essentially upon the charge carried by the ions in a gas, let us confine our attention for the present to neutral gases. Any conclusions deducible from a study of neutral gases can be immediately extended to the motions of ions in a gas, for the ions will behave very much like one neutral gas of very low concentration mixed with another neutral gas of much higher concentration.

Suppose that two gases at the same temperature and pressure, for example, hydrogen and carbon dioxide, are contained in separate compartments of a vessel, as shown in Fig. 7-6. Then

let the partition PO in the connecting tube be carefully removed without mechanically mixing the gases. Because of the fact that the CO_2 molecule is 22 times as heavy as the H_2 molecule, it might be assumed that the gases would remain separated. However, an experimental test will show that in a relatively short time the two gases will have intermingled so thoroughly that no difference in the concentrations of the two gases is detectable in any part of the vessel. This experiment is a good qualitative demonstration of the effect of the thermal agitation of the molecules in producing a homogeneous mixture.

There is another, much simpler, experiment with which everyone is familiar and which also illustrates the rapidity with which one gas diffuses through another. If a bottle of hydrogen sulphide is uncorked at some point in a room, in a very short time the odor of the gas can be detected at any point in the room—some of the molecules have found their way to every part of it. Here, again, if the air is not circulating, it is simply the thermal motions of the molecules that make this diffusion possible.

From experiment, we know that the net number of molecules of type 1 (for example, CO_2 in Fig. 7-6) that cross a unit area is proportional to the gradient of the concentration of molecules of type 1 in the direction perpendicular to the unit area. Thus, referring to Fig. 7-6 again,

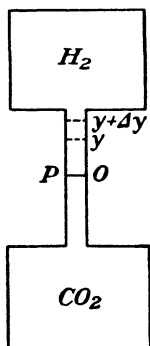


FIG. 7-6.

$$\eta_1 = -D_{12} \frac{dN_1}{dy} \quad (7-18)$$

Here η_1 is the net number of CO_2 molecules diffusing across a unit surface perpendicular to the connecting tube at P per second, N_1 is the concentration of CO_2 molecules at any point in the tube, dN_1/dy is the concentration gradient at P , and D_{12} is the coefficient of diffusion for molecules of type 1 (CO_2) diffusing into molecules of type 2 (H_2). The minus sign enters because there is a net movement of CO_2 upward only when the concentration of CO_2 is less in the upper part of the tube than in the lower.

It is the concentration gradient dN_1/dy that is the cause of the diffusion that takes place. If dN_1/dy were zero, the concentration of CO_2 would be constant throughout the connecting tube and just as many molecules would move downward across

a plane at P as move upward. But if the concentration of CO_2 is larger below P than above P , more molecules will move upward across a unit plane at P than move downward and a net diffusion of CO_2 upward will take place.

Immediately after the removal of the partition separating the CO_2 and H_2 in Fig. 7-6, the CO_2 molecules will start to diffuse upward and the H_2 molecules will start to move downward. Until a steady state is reached, there will be continuous change in the gas concentration of either kind of gas at any point in the container. To find the law governing this change, suppose we let η_1 represent the net number of molecules of CO_2 moving upward through a unit plane at y per second. Then, since η_1 is in general a function of y , the net number of molecules of CO_2 moving upward through a unit plane at $y + \Delta y$ per second will be $\left(\eta_1 + \frac{\partial \eta_1}{\partial y} \Delta y\right)$. The rate of increase of CO_2 molecules in a slab of thickness Δy is thus the number diffusing into the slab minus the number diffusing out of the slab, or

$$\frac{\partial}{\partial t} (N_1 \Delta y) = \eta_1 - \left(\eta_1 + \frac{\partial \eta_1}{\partial y} \Delta y\right)$$

Simplifying this, we have

$$\frac{\partial N_1}{\partial t} = - \frac{\partial \eta_1}{\partial y}$$

But, by the definition of diffusion,

$$\eta_1 = -D \frac{dN_1}{dy}$$

Therefore,

$$\frac{\partial N_1}{\partial t} = D \frac{\partial^2 N_1}{\partial y^2} \quad (7-19)$$

A solution of this equation would give the concentration of CO_2 molecules at any time and at any point in the connecting tube.¹ This equation is similar to a well-known equation in the theory of heat flow, and solutions can be found for it in particular cases.

¹ Further analysis shows that, in the general case of diffusion in three dimensions, the rate of change of concentration with time is given by

$$\frac{\partial N_1}{\partial t} = D \left(\frac{\partial^2 N_1}{\partial x^2} + \frac{\partial^2 N_1}{\partial y^2} + \frac{\partial^2 N_1}{\partial z^2} \right)$$

In most problems in conduction in gases, we shall be interested in conditions after they have reached a steady state. From Eq. (7-19) we can see, then, that in the steady state, when $\partial N_1/\partial t = 0$, the concentration gradient will be a constant and that the concentration at any point can be expressed in terms of the concentration at some reference point as follows:

$$N_1 = N_{10} + \frac{dN_1}{dy} y \quad (7-20)$$

where N_{10} is the concentration at $y = 0$.

If we assume that a diffusion process has been going on long enough for a steady state to be reached, it is easy to derive an expression for the diffusion coefficient, particularly if the diffusing

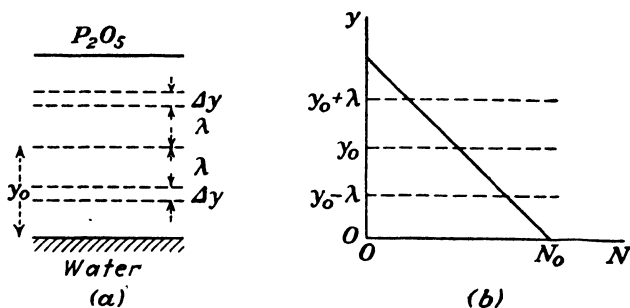


FIG. 7-7.—Unidirectional diffusion.

molecules have a very low concentration in comparison with the other gas present.¹ Such a situation might be encountered, for example, if we were considering the diffusion of water molecules from a water surface to some completely absorbing surface such as phosphorus pentoxide (see Fig. 7-7a). After a steady state has been reached, dN_1/dy is constant and the concentration of H_2O molecules varies uniformly from the water surface to the phosphorus pentoxide according to the linear relation given in Eq. (7-20) and as shown in Fig. 7-7b. In Fig. 7-7b the abscissas represent the molecular concentration at the various distances above the water surface.

Considering only the average motions of the molecules (*i.e.*, considering all the molecules as having velocities equal to the average velocity and free paths equal to the mean free path), we can try to compute an average value for the diffusion coefficient

¹ General reference 7-1, pp. 256-258.

cient D . Dropping the subscript 1, we see that at a point located a distance equal to the mean free path below y_0 the number of H_2O molecules in a slab of unit cross-sectional area and thickness Δy is $\Delta y[N_0 + (y_0 - \lambda) dN/dy]$. Similarly at a point a distance equal to the mean free path above y_0 the number of H_2O molecules in a slab of unit cross-sectional area and thickness Δy is $\Delta y[N_0 + (y_0 + \lambda) dN/dy]$. Now, on the average, one-sixth of the molecules in each of these slabs is moving so that in their next free path they will cross the plane at y_0 . To get the net number of molecules crossing at y_0 , we notice first that since the molecules have an average velocity $\bar{c}$, we can substitute $\bar{c} \Delta t$ for Δy in the expressions above. Then the number of molecules passing from Δy at $(y_0 + \lambda)$ down through the plane at y_0 in time Δt is $\frac{1}{6}[N_0 + (y_0 + \lambda) dN/dy]\bar{c} \Delta t$, and the number of molecules from Δy at $(y_0 - \lambda)$ passing up through the plane at y_0 in time Δt is $\frac{1}{6}[N_0 + (y_0 - \lambda) dN/dy]\bar{c} \Delta t$. Thus the net transfer of molecules through the plane at y_0 in a time Δt as a result of the concentration gradient is the difference of these two quantities, namely,

$$\begin{aligned} \frac{1}{6} \left[N_0 + (y_0 - \lambda) \frac{dN}{dy} \right] \bar{c} \Delta t - \frac{1}{6} \left[N_0 + (y_0 + \lambda) \frac{dN}{dy} \right] \bar{c} \Delta t \\ = -\frac{1}{3} \lambda \bar{c} \frac{dN}{dy} \Delta t \quad (7-21) \end{aligned}$$

The net number of molecules moving upward through y_0 in unit time is thus

$$\eta = -\frac{1}{3} \lambda \bar{c} \frac{dN}{dy} \quad (7-22)$$

But by our definition of diffusion we also have

$$\eta = -D \frac{dN}{dy} \quad (7-18)$$

Thus, the coefficient of diffusion in this case is given by

$$D = \frac{1}{3} \lambda \bar{c} \quad (7-23)$$

Although the derivation above is based simply on the average motion of the molecules, a more rigorous analysis of the problem yields the same expression for the coefficient of diffusion. It can be shown without too much difficulty¹ that in the general case of

¹ General reference 7-1, pp. 260-264.

two gases interdiffusing the diffusion coefficient is given by

$$D_{12} = \frac{\lambda_1 \bar{c}_1 N_2 + \lambda_2 \bar{c}_2 N_1}{3(N_1 + N_2)} \quad (7-24)$$

In this equation, N_1 and N_2 are the concentrations of the two kinds of molecule, $\bar{c}_1$, and $\bar{c}_2$ are the average thermal velocities of the two kinds of molecule, and λ_1 and λ_2 are the mean free paths of the two types of molecule, based on collisions of one type of molecule with the other.

In problems dealing with conduction in gases, the concentration of ions is usually negligible in comparison with the concentration of neutral molecules, and the general form for the coefficient of diffusion thus simplifies to the form given in Eq. (7-23). In this case, of course, λ refers to the mean free path of ions ending in a collision with neutral molecules.

Some idea of the magnitude of D for two gases interdiffusing and for ions diffusing through their own gas can be gained from the following table of experimental values.¹ All values in this table correspond to standard conditions of temperature and pressure. In the left half of the table, the column headed D gives values for the coefficient of diffusion of the first-named gas into the second, and in the right half of the table the columns headed D_+ and D_- give respectively the coefficients of diffusion of positive and negative ions in the gases named.

TABLE 7-1

Gas	D sq. m./sec.	Gas	D_+ sq. m./sec.	D_- sq. m./sec.
$O_2 \rightarrow H_2$	7.22×10^{-5}	Air	0.28×10^{-5}	0.43×10^{-5}
$H_2 \rightarrow O_2$	6.77×10^{-5}	O_2	0.25×10^{-5}	0.40×10^{-5}
$H_2 \rightarrow CO_2$	5.38×10^{-5}	CO_2	0.23×10^{-5}	0.26×10^{-5}
$C_2H_5OH \rightarrow$ air.....	1.02×10^{-5}	H_2	1.23×10^{-5}	1.90×10^{-5}
$CO_2 \rightarrow N_2O$	1.5×10^{-5}	N_2	0.29×10^{-5}	0.41×10^{-5}

It appears from the table that gases diffuse very much more quickly than ions and that negative ions diffuse slightly faster than positive ions. One cause of the slow diffusion of ions is that the charged ion forms a nucleus around which the molecules

¹ Derived from general reference 7-1, p. 256; J. S. TOWNSEND, *Phil. Trans.*, A195, 259, 1900; E. SALLES, *Ann. Physik*, 9th ser., 2, 273, 1914.

of the gas condense just as ions serve as a nucleus for the precipitation of water droplets. When this happens, a complex system is formed which diffuses slowly. The higher mobility of negative ions is a characteristic of most gases and explains why in certain cases ionized gas originally containing as many negative as positive ions eventually acquires a net positive charge. Thus, for example, if an electrically neutral but ionized gas is blown through a metal tube, the gas emerging will be positively electrified, because in its passage through the tube it has lost more negative than positive ions.

The differences between the experimental and theoretical values for the diffusion coefficient of gases can usually be explained as due to the variation of the mean free path with velocity and to what has been called the persistence of velocities. The latter effect was mentioned in the last section under the discussion of mobility; the term implies that the collisions between molecules are not completely elastic—more molecules are scattered in the forward direction than in any other. The factor by which the kinetic theory value of D as given in Eq. (7-23) or (7-24) must be multiplied in order to bring it in line with experimental values is not, in general, greater than 1.5; therefore, for most practical purposes the kinetic-theory value of D is sufficiently accurate.

The dependence of the coefficient of diffusion on temperature and pressure is evident from Eq. (7-23) above. Since $\bar{c}$ is proportional to the square root of the absolute temperature and λ is proportional to the absolute temperature and inversely proportional to the pressure, the diffusion coefficient will vary as the three-halves power of the absolute temperature and inversely as the pressure.

To illustrate the application of the principles of diffusion we can solve a simple problem.¹ Consider a discharge in which diffusion alone is the deionizing agent and uniform ionization by X rays is the ionizing agent. This is one of the simplest problems, but it will illustrate the general mode of attack on problems involving diffusion.

For simplicity, consider only a one-dimensional case. Thus, suppose the gas is confined between two metal plates A and B and that X rays shine on the gas and produce Q ion pairs per unit volume per unit time throughout the region. Diffusion is to occur only in the direction of the $+x$ -axis or $-x$ -axis (i.e. toward A and B). Let N_1 be the number of positive

¹ SLEPIAN, J., "Conduction of Electricity in Gases," pp. 45-46, Westinghouse Electric & Manufacturing Co., New York, 1933.

ions per unit volume and N_2 the number of negative ions per unit volume at any time t . Let D_1 be the coefficient of diffusion for positive ions and D_2 the coefficient of diffusion for negative ions. From the definition of D , we have

$$\eta = -D \frac{dN}{dx} \quad (7-18)$$

Consider a volume Δx m. thick cut from a cylinder of unit cross-sectional area extending between A and B (see Fig. 7-8). The net rate of increase of positive ionization in this slab per unit volume due to diffusion is given by Eq. (7-19) as

$$\frac{\partial N_1}{\partial t} = D_1 \frac{\partial^2 N_1}{\partial x^2} \quad (7-25)$$

An equation exactly similar to this except for a change of subscript would hold for negative ions. Because of their greater velocity and longer free path, negative ions diffuse out of a volume faster than positive ions, thus leaving behind a positive space charge. The resulting electric field tends to slow down the diffusion of negative ions and to accelerate the motion of the

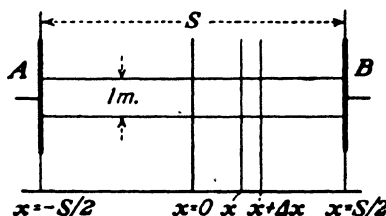


FIG. 7-8.

positive ions. In the end, the densities of positive and negative ions are almost equal, differing just enough to produce an electric field that just balances the difference in diffusion coefficients so that equal numbers of positive and negative ions are lost at the boundaries. Thus we can take an average diffusion coefficient D and apply it to both kinds of ion. Its value will

differ very little from D_1 . Since N_1 and N_2 differ by very little, an average value for these can also be taken. Call it N . The net rate of increase of ionization (positive or negative) in the slab per unit volume due to diffusion is thus

$$\left[\frac{\partial N}{\partial t} \right]_{\text{diffusion}} = D \frac{\partial^2 N}{\partial x^2} \quad (7-26)$$

Since the X rays produce equal numbers of positive and negative ions, the total gain in ionization per unit volume due to X rays is

$$\left[\frac{\partial N}{\partial t} \right]_{\text{X rays}} = Q \quad (7-27)$$

Upon combining Eqs. (7-26) and (7-27) the total rate of change in ion density is found to be

$$\frac{\partial N}{\partial t} = Q + D \frac{\partial^2 N}{\partial x^2} \quad (7-28)$$

If this equation can be solved, the ion density at any time and in any part of the region between A and B will be determined. For simplicity, consider

only the steady-state case when $\partial N/\partial t = 0$. For this case the rate of gain in ionization per unit volume is equal to the rate of loss of ionization per unit volume. Furthermore, since the only variables are N and x , total derivatives can be substituted for the partial derivative. Then,

$$\frac{d^2N}{dx^2} = -\frac{Q}{D} \quad (7-29)$$

If x is measured from the mid-point between A and B , then the conditions to be satisfied by N are $N = 0$ when $x = \pm S/2$, and at the boundaries the rate of loss of ions in the x -direction is one-half the total production. Thus, when $x = +S/2$,

$$\eta = -D \frac{dN}{dx} = \frac{QS}{2}$$

Integrating Eq. (7-29), we obtain

$$\frac{dN}{dx} = -\frac{Q}{D}x + C$$

After evaluating C ,

$$\frac{dN}{dx} = -\frac{Q}{D}x \quad (7-30)$$

Integrating again,

$$N = -\frac{Qx^2}{2D} + C'$$

After evaluating C' ,

$$N = -\frac{Qx^2}{2D} + \frac{QS^2}{8D}$$

or

$$N = \frac{Q}{2D} \left(\frac{S^2}{4} - x^2 \right) \quad (7-31)$$

From this, it can be seen that the maximum density occurs at $x = 0$ or at the mid-point between the boundaries. The value of the density here is

$$N_{\max} = \frac{QS^2}{8D}$$

The total number of ions in a cylinder of unit cross-sectional area extending between the walls can be found by integrating Eq. (7-31) between the limits $x = -\frac{S}{2}$ and $x = +\frac{S}{2}$. Thus,

$$\begin{aligned} n_{\text{total}} &= \int_{-\frac{S}{2}}^{+\frac{S}{2}} \frac{Q}{2D} \left(\frac{S^2}{4} - x^2 \right) dx \\ &= \frac{QS^3}{12D} \end{aligned} \quad (7-32)$$

All the ions generated in this cylinder are lost at the walls. Hence, the total loss in ionization per second is QS . We see, then, that although the

loss in ionization increases as the first power of the separation of the walls, the total ionization present increases as the cube of the distance between the walls. The ratio of these quantities, or the relative loss in ionization for a given total ionization, is thus inversely proportional to the square of the wall separation as shown in Eq. (7-33).

$$\frac{\text{Total loss in ionization per second}}{\text{Total ionization present}} = \frac{QS}{QS^3/12D} = \frac{12D}{S^2} \quad (7-33)$$

This means that, if an equilibrium condition is to exist in which the total number of ions between the walls is fixed, as it is, approximately, by the current flowing parallel to the walls, the number of ions generated per cubic meter throughout the volume by X rays—or by any other means—must increase rapidly as the separation between the walls is decreased. This is evident if we realize that at equilibrium just as many ions are lost to the walls as are generated by all ionizing agents acting.

An expression similar to Eq. (7-33) can be worked out for the distribution of ions over the cross section of a circular tube. From this, it follows at once that, the smaller the tube, the larger the proportionate loss of ions will be and hence the stronger the ionizing agent must be. It would be expected, then, that in order to maintain a given conductivity for a discharge through a small tube the electric field along the tube would have to be made larger and larger as the tube diameter is decreased. This is found to be so experimentally (see Fig. 10-12).

7-3. Recombination in the Gas.—When a gas containing ions of each sign is left entirely to itself, the ion concentration continually decreases. This cannot be explained as due to diffusion alone; for in the case of diffusion the ions are not actually lost but are simply spread throughout a larger volume. In the case of an arc in air, for example, ions are certainly lost from the region of the arc core by diffusion. But if the arc is not in a confining vessel, this diffusion can occur only to the gas surrounding the arc. Experiments have shown, however, that a short distance away from the arc the gas has a complete lack of ionization. It is evident from an experiment of this sort that the positive and negative ions, after diffusing away from the arc, must be combining with one another to form neutral atoms or molecules. The process by means of which the negative and positive ions come together to form neutral molecules is called *recombination*.

From experiments such as that described above, it is known that recombination of positive and negative molecular ions can occur very easily, whereas recombination of positive ions and electrons is very unlikely. This can be explained as due to the fact that on the average the kinetic energies of the positive and

negative molecular ions are about the same. The masses of the two kinds of ion are also about the same. Thus, from Eq. (7-7), it is easy to see that under these conditions an ion can lose a considerable amount of energy in an elastic collision. The probability of the ion having a very low energy when in the neighborhood of an ion of opposite sign is therefore large, and consequently the probability of recombination is great. According to classical collision theory, if two charged particles possess an appreciable amount of kinetic energy due to their thermal motions, they should be accelerated toward one another and move by on hyperbolic orbits similar to those shown in Fig. 7-9*a*. Recombination involves either the transfer of an electron from the negative to the positive ion as they pass or the settling down

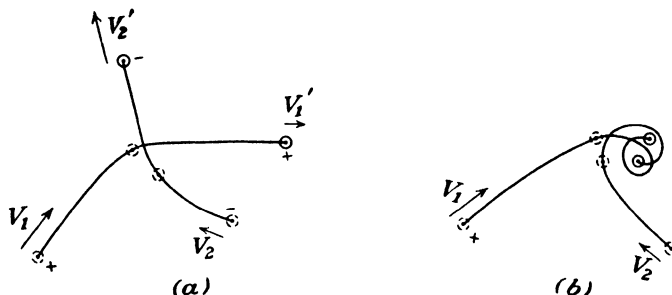


FIG. 7-9.—Ionic collision and recombination.

of the two ions into orbits about each other (Fig. 7-9*b*), until the electron has a chance to transfer itself from one to the other. The first of these processes is very improbable. In the first place the ions are in close proximity too short a time for the transfer to occur. In the second place, since the ions are charged, they suffer a deflection (or collision) while still a considerable distance apart, and consequently the forces holding the electron to the negative ion are usually much larger than the forces due to the passing positive ion. Recombination can occur, however, if one of the ions has a sufficiently small kinetic energy so that it can settle down into a stable orbit about the other ion until the electron has a chance to transfer itself.

It is considerations such as this which show why recombination between an electron and a positive ion is improbable. Referring again to Eq. (7-7), we see that the fraction of its energy which an electron can lose in collisions with neutral atoms is extremely

small in most cases. Thus, even in the extremely rare three-body collision in which the electron collides with a neutral molecule and the ion at the same time the chance of recombination is very slight, for the electron will still retain a large part of its original kinetic energy. In general, it will possess a sufficient amount of kinetic energy to overcome the attractive forces of the positive ion. Regardless of all this, it is still possible for recombination to occur between positive ions and electrons if the excess energy of the electron can be radiated. In a process such as this, the frequency of the emitted radiation would be given by the relation

$$h\nu = \frac{1}{2}mV^2 + q_e E_i \quad (7-34)$$

Here, $q_e E_i$ is a constant equal to the energy required to remove an electron from an atom,¹ and $\frac{1}{2}mV^2$ is the kinetic energy of the electron before recombination occurred. It is easy to see, then, that if this process occurred very frequently we should expect to observe a strong continuous spectrum in the discharge, since $\frac{1}{2}mV^2$ can have any value. Actually the amount of continuous spectrum observed in most discharges is very small, and for all practical purposes we can say that recombination does not occur between electrons and positive ions.²

In most discharges, electrons and positive ions recombine indirectly as a result of two processes. Electrons may attach themselves to neutral molecules to form negative molecular ions. Subsequently, a positive and a negative molecular ion may collide, and the electron may transfer itself from one ion to the other and thus form two neutral molecules. Apparently, the forces acting on an electron in the neighborhood of a neutral molecule vary much more rapidly than the inverse square of the separation. Thus an electron can approach very close to a neutral molecule before experiencing any force due to the nuclei and the orbital electrons. If, then, the electron has made a favorable series of

¹ See Sec. 8-1.

² What has been said here does not imply that recombination between electrons and positive ions *cannot* occur in any discharge. There is considerable evidence for the occurrence of this process in special low-pressure discharges where the probability of negative-ion formation is small. For example, at the center of an electrodeless discharge or in the negative glow of a low-pressure glow discharge, there is considerable spectroscopic evidence that recombination is occurring between electrons and positive ions. In most discharges, however, the magnitude of this effect is negligible.

collisions and its energy (or velocity) is sufficiently low, it can attach itself easily to the neutral molecule to form a negative ion. It has been shown experimentally that an electron must make many collisions before attachment can occur. Even so, the total number of collisions made per second may be so great that the electron will remain free only a fraction of a second. The rate of attachment of electrons to molecules at normal temperature and pressure has been expressed as an attachment constant A equal to the number of collisions made by the electron before becoming attached to a neutral molecule. The table below gives the value of A for several different gases as well as the number of electron impacts per second against gas molecules, Z_e , and the average life t in seconds of an electron at normal pressure and temperature.

TABLE 7-2*

Gas	A	Z_e	t
He, Ne, Ar { Hg, N ₂ , H ₂ }	∞		
CO	1.6×10^8	2.2×10^{11}	0.7×10^{-8}
Air	2×10^5	2.2×10^{11}	0.9×10^{-6}
O ₂	4×10^4	2.1×10^{11}	1.9×10^{-7}
H ₂ O	4×10^4	2.8×10^{11}	1.4×10^{-7}
Cl ₂	$< 2.1 \times 10^3$	4.5×10^{11}	$< 4.7 \times 10^{-9}$

* General reference 7-4, p. 193.

As can be seen from the table above, the average life of an electron ($t = A/Z_e$) in some gases is very short. In these gases, it is practically impossible to maintain a supply of free electrons. As the gas pressure is reduced, however, the electron mean free path increases; and, at pressures of the order of a millimeter, it may be impossible for the electron to make enough collisions to form a negative molecular ion before being lost through some other deionizing agent. Under these conditions, most of the negative ions will exist as free electrons. High gas temperatures, as we shall see later, also have the effect of increasing the average life of an electron and thus reducing the probability of recombination. With temperatures such as exist in the normal electric arc or in flames, it is likely that no negative ions are present. The fact that recombination proceeds slowly in gases

at high temperatures explains why the gas retains its conductivity for a long time. The ionization begins to disappear only as the gas cools. Because of this, it is possible for the hot gases from an arc (due to opening a switch, for instance) to drift between live parts relatively far away and cause a breakdown to occur. The low dielectric strength of the gas is due not to the high temperature but to the fact that the gas is ionized, and the high temperature merely prevents rapid loss of ionization.

The very short life of electrons in such gases as oxygen or water vapor explains why the probability of recombination between electrons and positive ions often seems to be unduly large. If even a very minute amount of oxygen or chlorine or a compound containing oxygen or chlorine is present in a gas such as helium, neon, or nitrogen, even though electrons do not attach themselves to the molecules of these gases, an appreciable amount of recombination will occur because the electrons can attach themselves so easily to oxygen and chlorine atoms. Because of the large effect that extremely small amounts of impurity can have, accurate determinations of the rate of attachment of electrons to neutral atoms are quite difficult if not impossible to make.

The rate of ion recombination has been experimentally found to be proportional to the concentration of positive ions and negative ions and is, therefore,

$$\frac{dN_+}{dt} = \frac{dN_-}{dt} = -\alpha_r N_+ N_- \quad (7-35)$$

The constant of proportionality α_r is called the *coefficient of recombination*. As in the case of the mobility and the coefficient of diffusion, this proportionality constant is really not a constant but depends on the temperature and the pressure of the gas as well as the velocities of the individual ions. Under any given set of conditions, however, it can be treated as a constant; and it is useful in this form for making estimates of the magnitude of recombination as a deionizing agent.

If the two kinds of ion are equally numerous, N per unit volume of each, Eq. (7-35) reduces to

$$\frac{dN}{dt} = -\alpha_r N^2 \quad (7-36)$$

That this equation should represent the rate of deionization is

evident if we consider that, in doubling the number of negative ions, we double the number of chances which each positive ion has of vanishing; and, in doubling both, we double both the number of ions which can vanish and the number of opportunities for recombination offered to each.

A direct method for the determination of α_r suggests itself after the integration of Eq. (7-36). If the ion density at any time t_0 is represented by N_0 , integration of the equation gives

$$\frac{1}{N} = \alpha_r(t - t_0) + \frac{1}{N_0} \quad (7-37)$$

Thus

$$\alpha_r = \frac{1}{(t - t_0)} \left(\frac{1}{N} - \frac{1}{N_0} \right) \quad (7-38)$$

By determining the ion density at various times after the removal of a uniform ionizing agent, it should be possible to determine the coefficient of recombination. The method employed in a recent measurement¹ was to use a shutter and synchronous switch arrangement. The gas between two uncharged parallel plates was irradiated for a short time by means of X rays; then a known interval of time after the X rays were cut off by means of the shutter, a strong electric field was applied between the plates and all the residual ionization drawn over to the electrodes. The total charge collected by the plates was measured by a Compton electrometer. Since the deflection of the electrometer was then proportional to the ion density at the instant the electric field was applied, it was easy to determine α_r by varying the time interval between the instant when the X rays were cut off and the instant of application of the electric field.

Table 7-3 gives typical values of the coefficient of recombination α_r at 760 mm. pressure and the temperature indicated.

Suppose we try to apply the kinetic theory of gases to the calculation of the coefficient of recombination, neglecting the force of attraction between oppositely charged ions and assuming that *every* collision between a negative and a positive ion results in recombination. Since we know by experiment that an ion must make many collisions before recombination occurs, we

¹ GARDNER, M. E., *Phys. Rev.*, **53**, 75, 1938.

should not necessarily expect to obtain a result that is smaller than the experimental value.

TABLE 7-3

Gas	α_r , cu. m./sec.	Temperature, deg. C.
Air*	1.23×10^{-12}	20
N ₂ *	1.06×10^{-12}	20
Ar*	1.06×10^{-12}	20
H ₂ *	0.28×10^{-12}	20
O ₂ †	2.08×10^{-12}	25

* LUHR, O., *Phys. Rev.*, **35**, 1394; **36**, 24, 1930. LUHR, O., and N. E. BRADBURY, *Phys. Rev.*, **37**, 998, 1931.

† GARDNER, M. E., *Phys. Rev.*, **53**, 75, 1938.

If the average number of collisions per second between one negative ion and a molecule is $\bar{c}/\lambda$ and the fraction of these collisions that occur with positive ions is N_+/N (N is the molecular concentration), we should expect the total number of collisions per second per unit volume between one negative ion and all the positive ions to be $N_+\bar{c}/N\lambda$. The total number of collisions per second per unit volume between all the negative ions and all the positive ions would then be $\bar{c}N_+N_-/\lambda N$; and since the assumption was made that every time a negative and a positive ion collide they recombine, this quantity would represent $\alpha_r N_+ N_-$. The value of α_r then becomes

$$\alpha_r = \frac{\bar{c}}{N\lambda} \quad (7-39)$$

Using this value of α_r , we calculate a value for the coefficient of recombination in oxygen at normal temperature and pressure of about 2×10^{-16} . The experimental value is 2×10^{-12} , or about 10,000 times greater than the value calculated above. It is obvious that the mutual electrostatic attraction between the particles is influencing their motions and causing recombination to occur at a much faster rate.

Although the electrostatic field around any ion extends to infinity, actually when the charged particles are some distance apart the change in velocity due to the electric field is so small in comparison with the total velocity that the effect of the electric field can be neglected. Thus, when the ions are some distance

apart, their thermal motions predominate and they may actually be carried away from one another. Thomson introduced the effect of mutual electrostatic forces by saying that, on the average, the motion of an ion is determined by ordinary kinetic-theory collisions until the ion happens to find itself so close to an ion of the opposite sign that its thermal energy is insufficient to carry it away. In other words, recombination will occur only when the ions are less than a distance r_0 apart, r_0 being defined by

$$\frac{q_e^2}{4\pi\epsilon r_0} = \frac{3}{2} kT \quad (7-40)$$

Thus, in Fig. 7-10, ion B has just made a collision with a neutral molecule and its average kinetic energy ($\frac{3}{2}kT$) is less than its potential energy with respect to ion A . The two will thus gradually approach one another and eventually recombine, although they may make several collisions with neutral atoms before doing so. The average kinetic energy of ion C , however, which has also just made a collision with a neutral molecule, is greater than its potential energy with respect to ion A ; and its thermal motion will, on the average, carry it away from A . In moving away, it may actually penetrate the sphere of radius r_0 ; but if it makes no collisions with neutral molecules inside the sphere, it will eventually escape. On this basis, Thomson¹ calculated the following expression for the coefficient of recombination:

$$\alpha_r = \pi r_0^2 (C_+^2 + C_-^2)^{\frac{1}{2}} (W_+ + W_- - W_+ W_-) \quad (7-41)$$

or

$$\alpha_r = \frac{q_e^4}{36\pi\epsilon^2 k^2 T^2} (C_+^2 + C_-^2)^{\frac{1}{2}} (W_+ + W_- - W_+ W_-) \quad (7-42)$$

Here, C_+ and C_- are the r.m.s. velocities of the ions, and the term involving W_+ and W_- is a probability term representing the chance that either ion will make a collision with a neutral molecule within the distance r_0 of the other and have its average

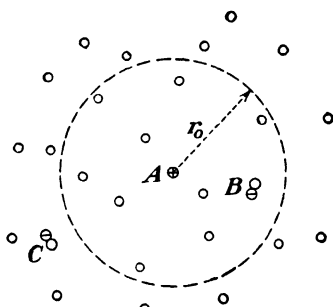


FIG. 7-10.

¹ General reference 7-3, p. 44.

kinetic energy reduced to $\frac{2}{3}kT$. W_+ represents the chance that a positive ion will collide with a neutral molecule while within a distance r_0 of a negative ion, W_- is the corresponding chance that a negative ion will collide with a neutral molecule while passing a positive ion, and W_+W_- is the chance that both will suffer collisions while within a distance r_0 of each other. This chance has already been included in both W_+ and W_- ; hence, it is necessary to subtract this term so as not to include this possibility

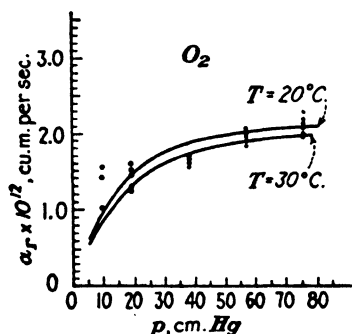


FIG. 7-11.—Dependence of recombination coefficient on pressure. (*M. E. Gardner, Phys. Rev.*, **53**, 75, 1938.)

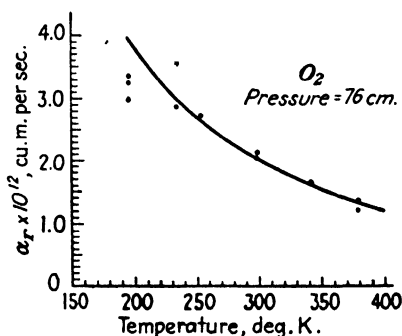


FIG. 7-12.—Dependence of recombination coefficient on temperature. (*M. E. Gardner, Phys. Rev.*, **53**, 75, 1938.)

more than once. These probability terms are expressed by an equation of the form

$$W = 1 - \frac{2}{x^2} [1 - e^{-x}(1 + x)] \quad (7-43)$$

where

$$x = \frac{2r_0}{\lambda} = \frac{q_e^2}{3\pi\epsilon kT\lambda} \quad (7-44)$$

The only difference between W_+ and W_- , then, occurs through the difference in mean free path of the two ions. In calculating W_+ , we should use the mean free path of the positive ions, and in calculation W_- , we should use the mean free path of the negative ions.

For pressures below atmospheric, Eq. (7-42) gives values of α_r that agree very closely with experiment. Figures 7-11 and 7-12 show how near the agreement is in the case of oxygen. In both curves the dots are experimental points and the solid lines are based on the theoretical equation. Figure 7-11 shows the

dependence on pressure, and Fig. 7-12 shows the dependence on temperature. As can easily be seen, the agreement is very good in both cases. In calculating the theoretical curves, it was necessary to assume that the ions each consisted of two molecules and consequently had a molecular weight of 64. This assumption has been verified by mass spectrograph measurements of the ionic mass. It was also necessary in the calculation to assume that the mean free path of the ions was one-fifth that of the neutral molecules. This too has been verified by experiment. The agreement between theory and experiment is striking considering the approximations that were made in the calculations.

Referring again to Eqs. (7-42), (7-43), and (7-44), we see that as the pressure increases the mean free path decreases and the term $(W_+ + W_- - W_+W_-)$ approaches unity. Thus, according to Thomson's theory, α_r should increase with the pressure but eventually should reach a constant value very much as indicated in Fig. 7-11. Experimentally, it has been found that, above atmospheric pressure, α_r actually decreases with an increase in pressure. This dependence on pressure has been predicted by Langevin¹ who assumed in his calculation of recombination that the ions continually move toward one another with a velocity determined by their mobilities and the electric field due to their charges. On this basis, he calculated a coefficient of recombination given by

$$\alpha_r = \frac{q_e}{\epsilon} (K_+ + K_-) \quad (7-45)$$

This theory, however, gives values of α_r that are much too large at pressures of the order of 1 atm. or less.

Loeb² has shown that if α_r is represented by a function of the form

$$\alpha_r = \alpha_{rT}f + \alpha_{rL}(1 - f) \quad (7-46)$$

it is possible to derive a curve for the dependence of α_r on pressure which agrees approximately with experiment over a wide range of pressures. In this equation, α_{rT} is the value of α_r derived from Thomson's theory, α_{rL} is the value derived from Langevin's theory, and $(1 - f)$ is a number representing the fraction of the negative ions formed, either by the ionizing agent or by electron

¹ LANGEVIN, P., *Ann. Chem. Phys.*, **28**, 287, 433, 1903.

² LOEB, L. B., *Phys. Rev.*, **51**, 1110, 1937.

attachment to neutral molecules, within a distance r_0 [see Eq. (7-40)] of a positive ion. The fraction f will naturally depend upon the pressure. Loeb's value for f is given by

$$f = e^{-ap^2} \quad (7-47)$$

where a is a constant generally determined by experiment. Using this value of f and Eq. (7-46), he derives the curve given in Fig. 7-13 which is in fair agreement with the experiments both at low pressures and at high pressures.

The dependence of α_r on temperature at atmospheric pressure is indicated in Fig. 7-12. The mathematical dependence of α_r on temperature can easily be estimated from Eq. (7-42). As the temperature is increased, x [Eq. (7-44)] decreases rapidly so that

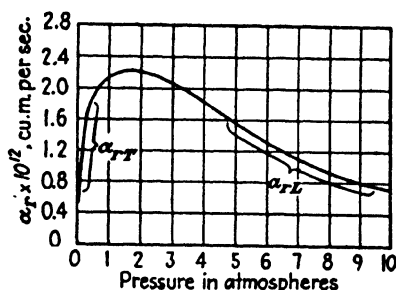


FIG. 7-13.—Variation of recombination coefficient with pressure. (L. B. Loeb, *Phys. Rev.*, **51**, 1110, 1937.)

at high temperatures the probability factor ($W_+ + W_- - W_+W_-$) can be approximated by $4x/3$. Then, since C is proportional to

$T^{1/2}$, α_r turns out to be proportional to $\left(\frac{1}{T^{1/2}} T^{1/2} \frac{1}{T^{1/2}}\right)$ or T^{-1} . The

coefficient of recombination thus depends strongly on the temperature, and at high temperatures there should be practically no recombination occurring. Tests with arcs at high pressure have revealed this to be the case. Consequently, in a discharge of this sort most of the ionization is lost by diffusion, and recombination occurs only in the cooler gas surrounding the arc.

7-4. Recombination at Walls and Electrodes.—From the argument of the preceding section, we realize that recombination between positive and negative ions is important only at relatively high pressures. At low pressures the mean free paths and translational kinetic energies of the free electrons are large, and

attachment of electrons to neutral molecules to form negative ions is quite improbable. For this reason, it is improbable that recombination between electrons and positive ions occurs. This means that the principal deionizing agents acting in the gas of a low-pressure discharge must be diffusion and the electric field. Recombination must occur, then, either on the walls of the enclosing vessel or on the electrodes.

In the case of insulated or insulating walls, electrons and positive ions may come up and condense on the wall. Since the electrons are so much lighter and diffuse so much faster than positive ions and since there is no way for the charge to leak off, the walls soon acquire a negative charge. This charge acts to repel other electrons and accelerate positive ions so that in the equilibrium condition just as many positive ions come up to the wall as electrons. At the surface, recombination occurs. The process involved in this recombination on the surface is not well understood. Apparently, electrons come up to the surface and adhere to it, giving up their kinetic energy in the form of heat. In this process the heat liberated per incident electron would be approximately¹

$$H_- = q_e \phi + \frac{1}{2} m C^2 - \frac{\eta_s}{\eta_p} \left(q_e \phi + \frac{1}{2} m C_s^2 \right) - \left(1 - \frac{\eta_s}{\eta_p} \right) 2kT \quad (7-48)$$

Here ϕ is the work function of the surface, C is the r.m.s. velocity of the incident electron, η_s is the number of secondary electrons emitted by η_p primary electrons, C_s is the r.m.s. velocity of the secondary electrons leaving the surface, and T is the difference in temperature of the surface and its external supports. In most problems the term involving $2kT$ is negligible in comparison with the other terms. Positive ions may condense on the surface in the same manner and there recombine with the electrons to form neutral molecules, recombination being possible directly between positive ions and electrons because both particles when on the surface have practically zero velocity and kinetic energy. In this process the energy liberated is

$$H_+ = q_e E_i - q_e \phi + L \quad (7-49)$$

where E_i is the energy required to ionize a neutral molecule, ϕ

¹ General reference 7-4, p. 182.

is the work function of the surface as before, and L is the latent heat of vaporization of the molecule. The kinetic energy of the incident ion does not appear in this equation because, in general, it is small enough to be neglected in comparison with the other terms. This energy would ordinarily also be converted into heat. If, instead of the positive ion condensing on the surface and recombining with the electron there, it pulls an electron from the surface as it approaches, owing to its electric field, and recombines with the electron just outside the surface, the energy liberated would be the same, although in this case a certain fraction of the energy of ionization $q_i E_i$ would probably be radiated as light. In either case the energy of the incident positive ion, simply by virtue of being ionized, is more than sufficient to extract an electron from the surface. That is, E_i is always greater than ϕ .

At a negative electrode, the process of recombination would be very similar to that at an insulated wall, the only difference being that in this case there is an initial excess of electrons on the surface and electrons from the gas are continuously repelled whereas positive ions are attracted. At a positive electrode, on the other hand, positive ions are repelled, but electrons are attracted. On striking the surface the latter may give up their energy to the electrode in the form of heat and pass directly into the metal to offset the initial deficiency of electrons.

Problems

1. The positive column of a certain discharge has a uniform voltage drop of 60 volts, and its length is 4 cm. The concentration of electrons in the discharge is 6.25×10^{20} per cu. m., and their mobility is 0.66 m./sec. per volt/m. If we could consider all the current in the discharge as due to the movement of electrons, what would be the current density in the discharge?

2. In a certain discharge tube in which the current consisted mainly of electrons the current density was 4,000 amp./sq. m. (0.4 amp./sq. cm.), and the voltage drop in the tube was 24 volts/m. Assuming that the density of electrons was constant throughout the tube and had a value of 1.5×10^{17} per cu. m., calculate the mobility of the electrons.

3. Calculate the mobility K of electrons through a gas of O₂ under standard conditions of temperature and pressure for a field strength of 10 volts/m. What is the average drift velocity of these electrons? **HINT:** Decide first whether this constitutes a high field or a low field; then base calculations on the electron mean free path as derived from the mean free path of oxygen molecules.

4. Assuming we have an arc in O_2 at atmospheric pressure, what would be the mobility of the electrons if the temperature of the arc is 5000° absolute and the field strength is 10 volts/m.? What is the average drift velocity of these electrons? HINT: Make use of the results of Prob. 3.

5. Calculate the mobility of electrons in O_2 under standard conditions of temperature and pressure when the field strength is 10,000 volts/m. What is the average drift velocity of these electrons?

6. Calculate the average velocity of drift of O_2 ions in oxygen at atmospheric pressure and $0^\circ C$. (a) when the field strength is 10^3 volts/m.; (b) when the field strength is 10^7 volts/m.

7. In an oxygen discharge under standard conditions of temperature and pressure, the average concentration of electrons and positive ions is the same and the field strength is 10 volts/m. What proportion of the total current is carried by electrons? By positive ions?

8. With reference to Prob. 7, what proportion of the total current is carried by electrons if the pressure is reduced to 10^{-3} atm. but the temperature remains the same? HINT: Use simplified expressions for high or low field as the case may be.

9. Under the conditions of Prob. 8, calculate the terminal energy of the electrons and positive ions. How does this energy compare with the average thermal energy of the gas molecules?

10. Under the conditions of Prob. 8, what fraction of the energy of the electron is lost in an average collision? What fraction of the energy of the positive ion is lost in each collision?

11. A plane surface of P_2O_5 is situated 3 cm. above and parallel to a water surface. Assume that the water molecules evaporate from the water surface and diffuse upward through air until they reach the P_2O_5 where they are all absorbed. If the concentration of water molecules close to the water surface is 5.7×10^{23} per cu. m. and the diffusion coefficient for water through air is 0.16×10^{-4} , how many kilograms of water are absorbed per second on 1 sq. m. of the P_2O_5 surface?

12. If the ion concentration in the neighborhood of an arc in nitrogen at atmospheric pressure and $0^\circ C$. is given by

$$N = N_0 \log \frac{1}{r}$$

where N_0 has the value $N_0 = 2 \times 10^{18}$ per cu. m., and r is a distance measured from the axis of the arc, how many ions are lost from the discharge per unit length, assuming all the loss is due to diffusion? HINT: Compute the number of ions diffusing out through a cylinder of radius r .

13. Compute the coefficient of diffusion of neon positive ions under standard conditions of temperature and pressure.

14. Compute the coefficient of diffusion of helium positive ions through helium at $100^\circ C$. and a pressure of 1 mm. Hg.

15. What would be the coefficient of diffusion of electrons under the conditions of Prob. 14?

16. If the concentration of positive ions is 3×10^{16} per cu. m., the concentration of negative ions is 2×10^{14} per cu. m., and the value of the

recombination coefficient is 2×10^{-12} , find the rate of loss of ionization from the gas.

17. Assuming the number of collisions made by an electron before attachment to form negative ions to be a constant, calculate the average life of an electron in O_2 under the conditions given in Prob. 4. HINT: Use the data of Table 7-2.

18. Compare the average life of an electron in O_2 with the time necessary for the electron to travel from the cathode to the anode in a discharge under standard conditions, if the anode-cathode separation is 1 cm. and the potential gradient is 10 volts/m. On the basis of the result of this problem, what should you say about the importance of recombination as a deionizing agent in this case?

19. Compute the coefficient of recombination for oxygen at standard conditions of temperature and pressure. Assume that the oxygen ion consists of two molecules and that the mean free path of the oxygen ion is one-fifth the mean free path of oxygen molecules. HINT: Use Eq. (7-42).

20. Compute the coefficient of recombination in oxygen at atmospheric pressure and a temperature of $1000^\circ C.$, the other conditions being the same as in Prob. 19. How does this value compare with that obtained in Prob. 19?

21. Compute the coefficient of recombination in oxygen at $0^\circ C.$ and a pressure of 10 atm., the other conditions being the same as in Prob. 19. Assume also that the mobility of positive ions is the same as the mobility of negative ions and that both correspond to the low-field condition. How does this result compare with the value obtained in Prob. 19?

General References

- 7-1. LOEB, L. B.: "Kinetic Theory of Gases," 2d ed., pp. 46-48, 252-276, Chap. XI, McGraw-Hill Book Company, Inc., New York, 1934.
- 7-2. DOW, W. G.: "Fundamentals of Engineering Electronics," pp. 435-444, John Wiley & Sons, Inc., New York, 1937.
- 7-3. THOMSON, J. J., and G. P. THOMSON: "Conduction of Electricity through Gases," 3d ed., Vol. I, pp. 20-84, Chap. III, Cambridge University Press, London, 1928.
- 7-4. COMPTON, K. T., and I. LANGMUIR: "Electrical Discharges in Gases," *Rev. Mod. Phys.*, **2**, 182-204, 217-221, 234-237, 1930.

CHAPTER VIII

PRODUCTION OF IONS IN THE GAS¹

8-1. Ionization Energies and Photoelectric Ionization in the Gas.—Although, in a high-vacuum tube, any current that flows must be due to the movement of charges liberated from the electrodes, it is possible, in a gas-filled tube, for a current of charges to be produced in the gas itself. One process of volume ionization that has already been mentioned in Sec. 7-2 is photoionization. Just as light shining on a surface will cause an emission of electrons to occur if the frequency of the light exceeds a certain minimum value, so light shining through a gas may cause the molecules of the gas to emit electrons if the frequency of the light exceeds a threshold value. The threshold value for gas ionization, however, is usually considerably higher than the photoelectric frequency limit of a solid substance.

The smallest energy required to remove an electron from a molecule or atom is called the *ionization energy* of that molecule or atom. It is customary to express this energy in terms of electron volts. Under these conditions the least energy required to produce ionization in the neutral atom or molecule is called the *ionization potential*. It is obvious that the ionization potential of a gas is similar to the net work function previously described for surfaces (see Sec. 6-3).

Table 8-1 gives the ionization potential E_i of the vapors of some of the more common elements and, for the sake of comparison, the photoelectric work function of some of the elements in the solid form. It is obvious from this table that, even in the case of the alkali metals, the ionization potential exceeds the work function by several volts.

Even cesium has an ionization potential that is only slightly lower than the work function of tungsten. Some idea of the significance of this can be gained by computing the low-frequency limit for ionization in cesium vapor. Since the threshold fre-

¹ General references for Chap. VIII will be found on p. 286.

quency is defined by the equation

$$h\nu_0 = q_e E_i \quad (8-1)$$

and $E_i = 3.86$ volts, we see that in the case of cesium ν_0 is 9.33×10^{14} sec.⁻¹. This corresponds to a wave length of 3,210 Å., well in the ultraviolet region of the spectrum. In the case of any other element the wave length must be even shorter. Since glass does not transmit light with a wave length shorter than about 2,800 Å. and even quartz will not transmit radiation shorter

TABLE 8-1

Element	First critical potential of vapor ^{a,b} E_c	Ionization potential of neutral vapor ^{a,b,c} E_i	* Second ionization potential of vapor ^{b,c} E'_i	Work function of solid ^{d,e}
Argon.....	11.57*	15.69	27.72	
Cadmium.....	3.78*	8.95	16.82	4.07
Cesium.....	1.4	3.86	23.4	1.67
Copper.....	1.39*	7.69	20.2	4.08
Helium.....	19.73*	24.48	54.16	
Hydrogen (atom).....	10.16	13.59		
Hydrogen (molecule).....	11.1	15.3		
Krypton.....	9.9*	13.9	26.4	
Lead.....	<2.6	7.38	14.96	3.75
Lithium.....	1.84	5.36	75.3	2.21
Mercury.....	4.65*	10.39	18.67	4.53
Neon.....	16.54*	21.47	40.89	
Nitrogen (atom).....	10.6	14.5	29.5	
Nitrogen (molecule).....	8.5	16.7		
Oxygen (atom).....	9.4	13.56	34.94	
Oxygen (molecule).....	6.1	14.1		
Potassium.....	1.61	4.33	31.68	1.77
Rubidium.....	1.53	4.16	27.3	1.82
Sodium.....	2.09	5.11	47.02	1.94
Tungsten.....		8.10		4.58
Xenon.....	8.3*	12.1	24.0	

* This critical potential corresponds closely to at least one metastable state.

^a "International Critical Tables," Vol. VI, pp. 69-73, McGraw-Hill Book Company, Inc., New York.

^b BACHER, R. F., and S. GOUDSMIT, "Atomic Energy States—As Derived from the Analyses of Optical Spectra," McGraw-Hill Book Company, Inc., New York, 1932.

^c "Smithsonian Physical Tables," 8th rev. ed., pp. 497-500, Smithsonian Institution, Washington, D. C., 1934.

^d Table 6-10, Sec. 6-12.

^e BECHER, J. A., *Bell System Tech. J.*, 14, 467, 1935.

than about $1,450 \text{ \AA}$,¹ it is obvious that, with most vapors, ordinary photoelectric ionization in which the light comes from an external source is not very important. Most of the photoelectric emission will take place at surfaces; for here the work function is very much smaller, and emission occurs much more easily than in the gas.

In the case of X rays or gamma rays shining on a gas, however, the wave length is less than $10 \text{ \AA}$. (the frequency is greater than $3 \times 10^{17} \text{ sec.}^{-1}$), and the energy of the radiation is large enough so that the rays can ionize the gas as well as produce surface emission. Because of the large energy of X rays, they are not appreciably absorbed in passing through a gas, and yet they can produce an appreciable amount of ionization in the gas. Consequently, X rays are often used when a uniform volume-source of ionization is desired.

In the table above are included the values for the so-called *second ionization potential* E'_i of the various vapors or gases. This second ionization potential corresponds to the amount of work required to remove a second electron from the atom or molecule when the molecule has already been ionized. From the table, it can be seen that these values are much higher in all cases than the corresponding values for the first ionization potential. Consequently, the probability of having doubly ionized molecules in a discharge is slight, although they are not entirely absent and in some highly ionized gases—intense discharges—they may be present in appreciable numbers.

It is a fundamental property of any atom or molecule that unless ionization occurs it can absorb only certain discrete amounts of energy. This energy has the effect of moving the most loosely held orbital electron or electrons in an atom to some larger orbit. Since the possible orbits in an atom or the possible energy states that an atom can assume are different for every chemical element, it is easy to see that the particular amounts of energy which can be absorbed by an atom have different values for the various elements. The average time that an atom exists in one of these so-called "excited states" is usually very short (of the order of 10^{-8} sec.). The tendency is for the atom to return to the state of lowest potential energy as soon as possible,

¹ Both glass and quartz begin to transmit again when the radiation has a wave length less than about 100 or 200 $\text{\AA}$.

emitting the extra energy in the form of light. Some atoms, however, possess excited states for which the average life is quite long—about 10^{-4} sec. Ordinarily, it is assumed, then, in theoretical discussions, that transitions in which radiation of energy occurs cannot take place from these long-lived, or *meta-stable*, states to the normal state of the atom. The only way for the metastable atoms to return to the normal state is to lose this excess energy in a collision with some other atom or molecule or with one of the walls or electrodes. Metastable states are encountered in many gases. For example, all the noble gases have important metastable states.

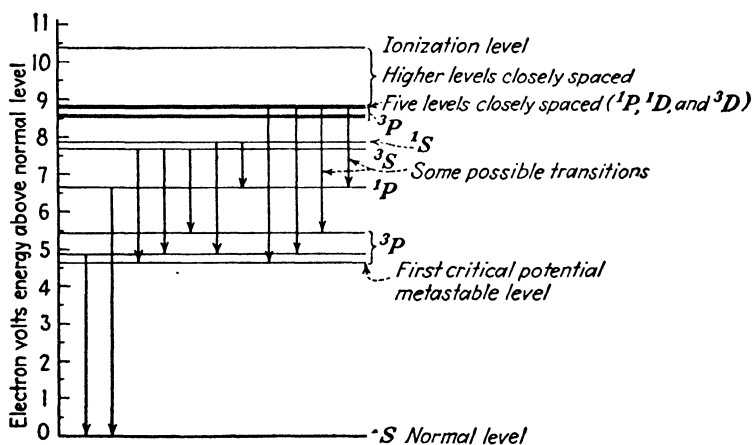


FIG. 8-1.—Energy level diagram for the mercury atom.

The smallest amount of energy that an atom or molecule can absorb, *i.e.*, the difference in the energy possessed by the atom or molecule in its most stable state and in its first excited state¹ is called the *first critical energy* for the atom or molecule. It is this energy which, expressed in electron volts, is called the *first critical potential* E_c (see Table 8-1 and Fig. 8-1). The highest excited state of an atom or molecule naturally corresponds to ionization or the complete removal of one orbital electron, and the energy required for this process is the ionization energy. From the table and from Fig. 8-1, it can be seen that there is a considerable spread between the first critical potential and the ionization potential. Because of this the number of transitions

¹ In general, this is a group of energy states differing, however, by only a fraction of an electron volt.

between excited states occurring in a gas is always considerably greater than the number of transitions resulting in ionization, and consequently a considerable emission of light usually accompanies a high rate of production of ionization.

The radiation or energy given off in a transition from the first excited state to the normal state may sometimes considerably alter the characteristics of a discharge as when, for example, the gas contains even a slight amount of an impurity that can be ionized by this radiation. A particular case might be a hydrogen discharge containing mercury vapor as an impurity. Since the first critical potential of molecular hydrogen is larger than the ionization potential of mercury vapor, it is possible for the light emitted by a hydrogen molecule in returning to its normal state to be absorbed by a mercury atom and result in ionization of the mercury. In this case, since the number of excitation transitions may be large in comparison with the number of ions produced in the hydrogen, it is possible for a large percentage of the mercury vapor to become ionized and quite materially to alter the characteristics of the discharge.

An example showing the effect of metastable atoms would be the case of a discharge in pure mercury vapor. Since the first excited state of mercury is a metastable state, atoms in this state will retain an energy of 4.65 volts for about 10^{-4} sec. This is a long enough time, on the average, for the atom to move over, strike, and give up its energy to one of the electrodes or walls; and if the work function of the electrode or wall is less than 4.65 volts, emission of an electron can take place. In discharges in the noble gases, ionization due to impacts of metastable atoms on the electrodes and walls is an important source of ionization and must be taken into account in the calculation of discharge characteristics.

8-2. Thermal Ionization.—Ionization of a gas can occur when the average energy of the molecules becomes so great that the energy transferred in a collision between two neutral molecules is sufficient to ionize one of them. Actually, however, ionization by this process occurs, to an appreciable extent, only in high-pressure discharges such as free arcs in air or in oil circuit breakers. At normal temperature (300°K.) the average kinetic energy of the gas molecules corresponds to less than 0.04 e.v. Consequently, at this temperature, the fraction of the total number

which have energies greater than that necessary for ionization of the gas is extremely small regardless of the gas used.

The problem of computing the amount of thermal ionization present at any particular temperature is very complex, for at high

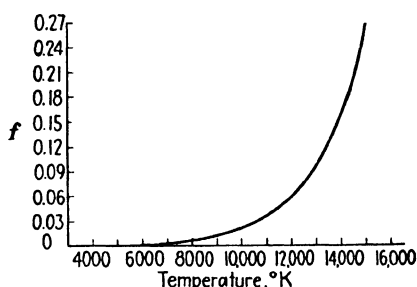


FIG. 8-2.—Fraction of molecules thermally ionized at any temperature.

temperatures it is probable that ionization does not occur simply owing to the thermal motions of the molecules. It is probable that at very high temperatures there is a considerable photoelectric emission from the walls and perhaps in the gas itself in addition to a considerable amount of ionization caused by the free electrons which are liberated by any of these processes. Experimental verification is also difficult, for it is not easy to obtain controllable temperatures of 5000 to 10,000°K.

Qualitative estimates of the amount of thermal ionization present at any temperature under equilibrium conditions have been made on the basis of the so-called "Saha" equation. One form of this equation is as follows:

$$\frac{f^2}{1-f^2} p = 2.4 \times 10^{-4} T^{\frac{5}{2}} e^{-\frac{q_i E_i}{kT}} \quad (8-2)$$

In this equation, f is the fraction of the total number of molecules that are ionized, p is the gas pressure expressed in millimeters of mercury, T is the gas temperature in degrees Kelvin, and E_i is the ionization potential of the gas. If f is small enough so that $(1 - f^2)$ can be set equal to 1, we see that the fraction of the total number of molecules that are ionized at any given temperature is proportional to $e^{-\frac{q_i E_i}{2kT}}$. Figure 8-2 is a plot of this curve for the

case of air at atmospheric pressure. The ionization potential in this case was assumed to be 15 volts. The very rapid increase in ionization with temperature is evident from the curve. It is

also evident that there is very little ionization below about 4000°K.

By means of the Saha equation, it is possible to estimate the rate of production of ions due to thermal ionization. Figure 8-3 represents the results obtained. Here again we see that the rate of production of ionization is very small until the temperature reaches a few thousand degrees Kelvin but that in this temperature range it is increasing rapidly.

8-3. Ionization by Collisions of Electrons with Gas Atoms.—

Probably the most important source of ionization throughout the body of a gas is the collisions of free electrons with neutral atoms or molecules. Because the mass of the electron is so very small, the energy that it loses in an elastic collision with a neutral molecule is only a small fraction of its total kinetic energy.

In Sec. 7-1 of the last chapter, we saw that this fraction is given by

$$f = 2.66 \frac{mM}{(m + M)^2} \left(1 - \frac{\bar{W}_M}{\bar{W}_m} \right) \quad (7-7)$$

where $\bar{W}_M$ and M are, respectively, the average energy and mass of the molecules and $\bar{W}_m$ and m are the average energy and mass of the electrons. This is approximately equal to m/M even when the average energy of the electrons is only 20 or 30 per cent larger than the average energy of the molecules. Thus, even with a comparatively small electric field, if the electron can make a sufficient number of collisions the average energy of the electron can become many times that of the neutral molecules. In fact, referring to Eq. (7-10) we see that the average energy of the electron is given approximately by

$$\bar{W}_m = \frac{qF\lambda_e}{2.31} \sqrt{\frac{M}{m}} \quad (8-3)$$

so that, even if the electric intensity were only 100 volts/m., the

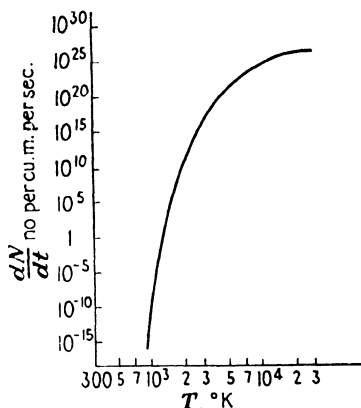


FIG. 8-3.—Dependence of rate of thermal ionization on temperature. (Derived from A. v. Engel and M. Steenbeck, "Elektrische Gasentladungen," Verlag Julius Springer, Berlin, 1932, Vol. 1, p. 90.)

average energy of electrons in oxygen would be about 4.5 e.v. at 1 mm. pressure. This corresponds to a temperature of about 35,000°K. Under conditions such as these, the actual kinetic energy of a fairly large fraction of the electrons would thus be larger than the ionization energy of the molecules, and ionization could occur owing to the collisions of these electrons with neutral molecules.

It is important to note that, even though we speak of ionization by collision between electrons and neutral molecules, the processes involved, as in thermal ionization, may be complex. The electrons may ionize directly in colliding with a neutral molecule; or one electron may excite a molecule, and a subsequent electron may ionize it. This would be called *ionization by successive impacts*. On the other hand, an electron might excite a neutral molecule which subsequently, in returning to its normal state, would give off radiation and cause photoionization either in the gas or at the walls of the discharge. There is no doubt that in any particular discharge a combination of all these processes may be taking place.

The probability of an electron ionizing a molecule in a collision is difficult to determine experimentally, but estimates of this probability can usually be obtained by projecting a stream of electrons of known kinetic energy into a gas and first determining the average number of *ionizing* collisions made per unit distance by the electrons and then dividing this number by the number of collisions that the electrons make in going unit distance. The latter is the harder to determine accurately, and it has been customary to use the average number of ionizing collisions per unit distance in the direction of the field as the measure of ionizing efficiency. As a matter of fact, this quantity is much more useful in the theory of conduction in gases than the actual probability of ionization.

Experimentally, it has been found that within the range encountered in most gas discharges *the number of ionizing collisions per meter made by an electron*, which we shall call α , is approximately directly proportional to the excess of energy of the electron above the ionizing energy. That is,

$$\alpha = C(E - E_i) \quad (8-4)$$

where E is the actual energy of the electron expressed in volts

and E_i is the ionization potential of the gas. This equation holds only when E is less than $2E_i$. In that range the constant C has the following values¹ for some of the commoner gases when the pressure is 1 mm. and the temperature is 0°C :

TABLE 8-2

Gas	C
Argon.....	71
Helium.....	4.6
Neon.....	5.6
Oxygen molecule.....	24
Nitrogen molecule.....	30
Hydrogen molecule.....	21
Air.....	26
Mercury.....	82.8

The ionizing efficiency of an electron does not continue to increase indefinitely as its energy is increased. The curve for α

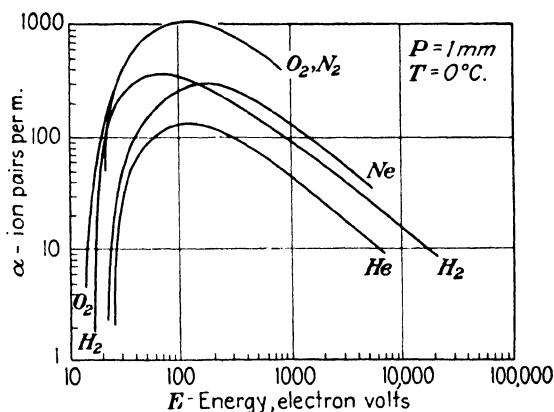


FIG. 8-4.—Ionizing ability of electrons in various gases as a function of electron energy. (Derived from A. v. Engel and M. Steenbeck, "Elektrische Gasentladungen," Verlag Julius Springer, Berlin, 1932, Vol. 1, p. 35.)

reaches a maximum when the electrons have an energy between 50 and 150 e.v. and then decreases with a further increase in the energy of the electrons. Typical curves of ionization efficiency for several gases are shown in Fig. 8-4. As the energy of the electron increases, it is possible for the electron to remove two, three, four, or even five orbital electrons from the neutral molecule in a single collision. Bleakney² has measured the probability

¹ General reference 8-4, p. 38.

² *Phys. Rev.*, **35**, 139, 1930.

of multiple ionization in mercury vapor when the primary electrons have energies below 400 volts. His results are represented in Fig. 8-5. In the curves of Fig. 8-4, no correction has been made for the possibility of multiple ionization. Hence, the curves represent the total ionization produced per meter by the electron, double or triple ionization being counted as two or three ions. All curves are reduced to 1 mm. pressure and 0°C .

The relationship of ionization by electron impacts to particular discharges will be discussed in more detail in Sec. 8-6. Before leaving this subject, however, it is worth emphasizing again

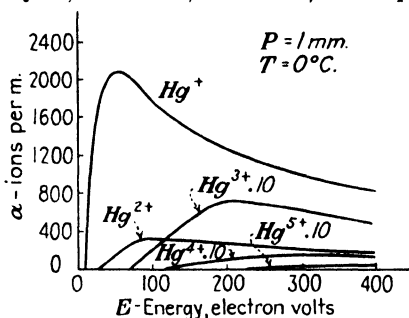


FIG. 8-5.—Ionizing ability of electrons in mercury vapor. (W. Bleakney, *Phys. Rev.*, **35**, 139, 1930.)

that the energy which an electron must have before it can produce an appreciable amount of ionization is always considerably larger than the ionizing potential. The ionization potential determines only the energy at which ionization by collision starts. The greatest amount of ionization by collision occurs when the electron has an energy between five and ten times that given by the ionization potential.

8-4. Ionization by Collision of Positive Ions with Gas Atoms.—

Positive ions are much less efficient than electrons in producing ionization by collision. Because of its large mass, a positive ion can lose a large part of its kinetic energy in an elastic collision; consequently, unless it is in an extremely high electric field, it never attains an energy much larger than that of the neutral gas molecules. Consequently, in most gas discharges, the number of positive ions that even attain an energy equal to the ionizing energy of the gas is only a very small fraction of the total number of positive ions present. Furthermore, whereas an electron begins to produce ionization by collision when it has

an energy equal to the ionizing energy of the gas, we know by experiment that a positive ion must have an energy of several hundred volts before it can begin to produce ionization by collision and approaches the efficiency of an electron only when it has an energy greater than 10,000 volts—*i.e.*, when its velocity is the same as that of an electron having an energy equal to the ionizing energy of the gas.

In view of all this, it is safe to say that secondary effects produced by positive ions are much more important than ionization by collision. These secondary effects can take the form of secondary emission of electrons from walls and electrodes or of inelastic collisions with gas atoms in which the light emitted causes photoelectric emission in the gas or at surfaces.

In any collision between a charged particle and a neutral molecule, ionization takes place because of the electric force exerted on the planetary electrons in the molecule. Thus, though fast-moving positive ions, such as alpha particles, produced in the disintegration of radioactive substances, in colliding with neutral molecules, may be expected to produce ionization in the same way as electrons, collisions between slow positive ions and neutral molecules, in general, are much more complex. Since usually the only difference between the two colliding particles is that one has a deficiency of planetary electrons, the collision may result in one of three different processes. A transfer of kinetic energy only may take place, an electron may be transferred from the molecule to the ion, or an electron may be ejected from the molecule. It is only this third process which alters the ion concentration, for then the result of the collision is an additional positive ion and electron. All these processes are known to occur.

The curves of ionizing efficiency of positive ions are similar in shape to those for electrons, the difference being that, in general, the ionizing efficiency is much lower at all velocities and the minimum energy for ionization is much higher. Table 8-3 gives the experimentally determined minimum ionizing energy in volts of some of the alkali positive ions in inert gases. Data for these ions and gases are easier to obtain than any other; for negative molecular ions are not formed in the noble gases, and it is easy to obtain a large supply of low-energy alkali ions.

TABLE 8-3*

Gas \ Ions	Li	Na	K	Rb	Cs
Ne	307	175	320	420	437
Ar	100	105	95	180	365
Kr	420	400	80	100	143
Xe	250	360	120	145	105

* General reference 8-1, Vol. II, p. 123.

8-5. The Townsend Discharge.—The electric field is one of the most important ionizing agents, as well as one of the most important deionizing agents, that may be present in any discharge. In the preceding sections, we have seen how it may act to produce ionization by collision. Now let us investigate the manner in which the electric field, through its effect on ionization by collision, may contribute to determining the characteristics of some typical discharges. Let us start with the so-called "Townsend discharge."

The name *Townsend discharge* is now more or less generally applied to that class of *non-self-maintaining* discharges characterized by the fact that the principal means of producing ionization is through ionization by collision, but excluding those discharges which, for their maintenance, depend on a thermionic cathode. For example, the Townsend discharge always occurs just before breakdown or sparking takes place in a spark gap. It also occurs in a gas-filled phototube when the tube voltage exceeds about 30 volts. The discharges are named for J. S. Townsend who as long ago as 1901 investigated current amplification through ionization by collision and proposed a theory for the results obtained. Although our present-day knowledge of collision processes has necessitated some changes and modifications in his theory, it is still qualitatively correct and is certainly much simpler and easier to comprehend than the more exact theories. The discussion and derivations that follow are all based on Townsend's original theory, some modifications being introduced to bring the theory more in line with our present understanding of the nature of conduction in gases.

Consider Fig. 8-6 which is the characteristic voltage-current curve of a gas-filled phototube and is similar to the voltage-current curve of any gas discharge when the applied voltage is

quite low. It has three distinct parts, a rising part OA , a horizontal part AB , and another rising part BC . The first two parts OA and AB are familiar and can be attributed to the same processes as occur in high-vacuum tubes. In the region OA , for example, the current may be limited by space charge, and all the negative ions may not be reaching the anode. This would be the case if the electrons were emitted from a thermionic cathode or if they were produced close to the cathode by X rays, or by radioactive substances. If the negative ions are produced throughout the space between cathode and anode by such means as X rays, cosmic rays, or the disintegration of radioactive sub-

stances or if they originate at the cathode owing to the photoelectric effect and hence leave the cathode with an appreciable velocity, the current is not apt to be limited by space charge but can be limited by the fact that the negative ions or electrons have appreciable velocities and diffuse away and recombine with positive ions on the walls or in the gas before they can reach

the anode. However, when the voltage reaches a certain value, the current may become practically constant, especially in low-pressure discharges, a fact indicating that all the ions are finding their way to either the anode or the cathode. This corresponds to the current-saturation region AB . As the potential between the electrodes is increased still more, the current again begins to increase and the region BC is reached in which a new mechanism of ionization begins to be noticed, namely, ionization by collision. From B onward, the rate of increase of current becomes more and more pronounced until finally it approaches infinity and the whole form of the discharge changes.

Let us now consider in more detail that region of the Townsend discharge represented by BC in the curve of Fig. 8-6. Assume that we have two plane-parallel electrodes and that ultraviolet light shines on the negative electrode (see Fig. 8-7). Thus, all the primary electrons are formed at the cathode surface, and all of them must pass through the whole thickness of gas before

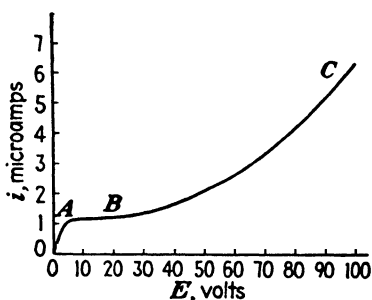


Fig. 8-6.—Voltage-current curve of a gas-filled phototube.

being lost at the anode. Let α represent the *number of ion pairs formed by an electron per meter of drift toward the anode*. The effect of the positive ions in producing ionization can be neglected for the present; for electrons, as seen above, are more efficient in producing ionization than positive ions are. Take the origin of coordinates on the cathode surface, and lay the x -axis perpendicular to the two plates. Thus, the cathode lies at $x = 0$, and the anode at $x = d$. Since α depends on the field strength and the pressure, these quantities will be assumed to be constant throughout the region between the two electrodes. This is not an unreasonable assumption, for we have already assumed that the potential difference between the plates is high enough to

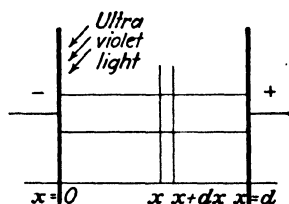


FIG. 8-7.

produce saturation. However, assuming F constant does mean that if d is varied the potential between the plates will have to be varied so that F can remain constant. Suppose that η_0 electrons are given off by unit area of the cathode per second and that η electrons per second are drifting across a unit area perpendicular to the x -axis at any point x between the two electrodes. Denote by $d\eta$ the number of ion pairs formed per second in a slab at x of unit cross section and thickness dx . Thus, the number of new electrons formed by η electrons traveling a distance dx is

$$d\eta = \eta\alpha dx \quad (8-5)$$

Integrating,

$$\eta = Ae^{\alpha x} \quad (8-6)$$

To evaluate the arbitrary constant A , we need only note that, when $x = 0$, $\eta = \eta_0$, the number of electrons liberated from the cathode per second. Thus, $A = \eta_0$ and

$$\eta = \eta_0 e^{\alpha x} \quad (8-7)$$

The current density at any point between the two electrodes will consist of electrons moving toward the anode and positive ions moving toward the cathode. But, at the anode, all the current consists of electrons, since no positive ions are liberated by the anode. Thus, the total current density between the plates is given simply by

$$J = q\eta_0 d \quad (8-8)$$

where η_d is the number of electrons striking unit area of the anode per second. It follows, then, that

$$J = q_e \eta_d e^{ad} \quad (8-9)$$

or

$$J = J_s e^{ad} \quad (8-10)$$

Here, J_s is the saturation current density that exists before ionization by collision sets in. This relation [Eq. (8-10)] has been verified over a wide range of field strength and pressure for many gases.

Three important facts can be deduced from Eq. (8-10):

1. A discharge in which "self-amplifying ionization" occurs depends for its maintenance on a continuous source of primary electrons. Removal of the ultraviolet light, in the example above, causes the discharge to cease immediately, because J_s would then be zero.

2. The current flowing will depend on the field strength. Since α depends on the field strength, an increase or decrease in α caused by a change in field intensity will produce a corresponding exponential change in the current.

3. The current flowing will increase exponentially with the separation between electrodes if the field intensity is maintained constant. This result follows immediately from Eq. (8-10), for α and J_s are constant. It can be explained qualitatively by the fact that an increased electrode separation means that a greater quantity of gas is enclosed between the electrodes and that consequently an electron will make more ionizing collisions in going from cathode to anode. It is true that the potential increases (if F is constant) with electrode separation; but this is only a linear increase and the current increase is an exponential function of the electrode separation.

The solid curve in Fig. 8-8 is a semilogarithmic plot of the variation of current with electrode spacing in a particular case in which F was 12,650 volts/m., the pressure was 0.973 mm. Hg,

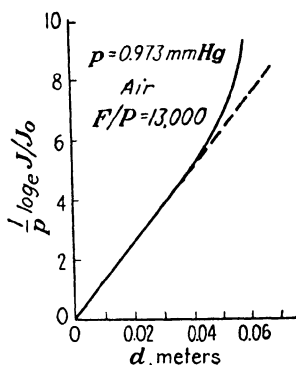


FIG. 8-8.—Dependence of current density on electrode separation. (F. G. Sanders, *Phys. Rev.*, **44**, 1020, 1933.)

and the gas was air. For values of d up to about 0.04 m. the curve is a straight line, a fact verifying the exponential relation in Eq. (8-10). From the slope of the straight portion of the curve, it is possible to deduce the value of α . In this case the computed value of α is 123.6. Thus, in the case of this discharge, each negative ion as it drifts toward the anode produces, on the average, 123.6 pairs of new ions in every meter.

The theory outlined above could be applied equally well to a case in which the primary source of electrons is a volume source of ionization such as X rays, instead of the surface source considered. In this case, also, we should expect to get a saturation current when the electric field is just sufficient to sweep out all the electrons that are generated by the X rays but not sufficient to produce an appreciable amount of ionization by collision. Now, however, the number of new ion pairs formed in a slab of unit cross section and thickness dx parallel to the electrodes by the X rays and the electric field is

$$d\eta = \eta\alpha dx + \eta_0 dx \quad (8-11)$$

where η_0 is the number of ion pairs formed per unit volume per second by the X rays and the other quantities have the same meaning as before. Integration of this equation yields the result that

$$J = J_0 \left(\frac{e^{\alpha d} - 1}{\alpha d} \right) \quad (8-12)$$

This equation is very similar to Eq. (8-10) and exhibits much the same properties.

Ionization by collisions of electrons with gas molecules is not sufficient to explain the current amplification that takes place in an actual discharge as the electrodes are moved farther and farther apart. As Fig. 8-8 shows, the current increases much more rapidly than can be explained by a simple exponential increase and, in fact, tends to increase without limit as a certain critical separation is approached. Ionization by electron impacts is adequate to explain the lower part of the curve of current vs. electrode separation; but where the actual curve begins to diverge from the exponential line, some new source of ionization must be gaining in importance.

The answer that occurred to Townsend was that positive ions, too, gain sufficient energy to produce ionization by collision.

On the basis of this hypothesis, he was able to explain the observed current increase. We know now, however, that this explanation is not right. Since positive ions must have an energy of about 100 volts before they can begin to produce ionizing impacts and since they rarely have an energy much larger than the thermal energy of the gas molecules, there is little likelihood of ionization by collision of positive ions with gas molecules being important for conduction except under unusual circumstances.

Townsend, however, suggested an alternative mechanism for the observed current increase that now seems to be a closer approximation to actual conditions than ionization by collision by positive ions. This can be stated as follows: If the positive ions do not ionize neutral gas atoms by collision, they may still possess enough energy to liberate secondary electrons when they strike the cathode. This seems to be a sufficient explanation for the observed current amplification in some discharges, especially at low pressures. At high pressures, this theory, too, fails to satisfy all the observed conditions.

There is probably no single mechanism that can explain the current increase under all conditions. Even at low pressures, there is probably an appreciable amount of ionization due to a third effect, namely, secondary photoelectric emission from the cathode caused by the light emitted following excitation collisions between electrons or positive ions and neutral molecules. A fourth mechanism is probably active with large electrode separations and at high pressures. Under these conditions, it is possible to explain the current amplifications simply on the basis of the ionizing collisions of electrons, for space charges certainly develop which distort the electric field and cause the ionizing ability of the electrons to vary widely in the various parts of the discharge.

Ignoring this last possibility for the present, let us derive an expression for the current between two electrodes in a gas when the auxiliary ionization is due to secondary emission from the cathode, the secondary emission being caused by the impact of positive ions or by radiation from the gas following excitation collisions between the particles in the gas or by some combination of these processes. Let η_p be the number of primary electrons emitted per unit area owing to the action of the external light

source on the cathode and let η_s be the number of secondary electrons emitted per unit area owing to any of the secondary ionizing processes mentioned above. Then, the total number emitted is

$$\eta_0 = \eta_p + \eta_s \quad (8-13)$$

The problem now is simply one of ionization by collision of electrons alone, and from Eq. (8-7) we know that the number of electrons crossing unit area of a plane at x parallel to the electrodes in the presence of a uniform electric field is

$$\eta = \eta_0 e^{\alpha x} \quad (8-7)$$

and the total number reaching the anode is

$$\eta_d = \eta_0 e^{\alpha d} \quad (8-14)$$

But η_0 of these come directly from the cathode; hence, the number of new ion pairs generated in the gas is

$$\eta_d - \eta_0 = \eta_0 (e^{\alpha d} - 1) \quad (8-15)$$

Now, let γ stand for the average number of secondary electrons emitted from the cathode for each new positive ion formed in the gas. Thus, the total number of secondaries formed on unit area of the cathode per second is

$$\eta_s = \gamma \eta_0 (e^{\alpha d} - 1) \quad (8-16)$$

Substituting this value of η_s in Eq. (8-13) and solving for η_0 , we find

$$\eta_0 = \frac{\eta_p}{1 + \gamma - \gamma e^{\alpha d}} \quad (8-17)$$

Therefore, Eq. (8-7) becomes

$$\eta = \frac{\eta_p e^{\alpha x}}{1 + \gamma - \gamma e^{\alpha d}} \quad (8-18)$$

and the number of electrons passing into the anode is, from Eq. (8-14),

$$\eta_d = \frac{\eta_p e^{\alpha d}}{1 + \gamma - \gamma e^{\alpha d}} \quad (8-19)$$

Since the saturation current, the maximum current possible under high-vacuum conditions, is determined by the number of electrons emitted from the cathode by the external light source, we have

for the total current density

$$J = J_s \frac{e^{\alpha d}}{1 + \gamma - \gamma e^{\alpha d}} \quad (8-20)$$

This equation reduces to Eq. (8-10) when γ is zero or when d is very small. Thus, we should predict an initial exponential increase in the current as shown in Fig. 8-8. As d increases, however, the denominator of Eq. (8-20) decreases and the current should increase faster than a simple exponential function. Using this equation, we can compute a value of γ to make the equation agree with experimental results. In the example cited above, we found α to be 123.6. The value of γ need be only 0.00061 in order to fit the upper curve in Fig. 8-8. This means that if only 0.00061 secondary electron is produced for each ion pair produced in the gas, or if only one electron is produced for about every 1,640 ion pairs produced in the gas, then the current predicted by Eq. (8-20) will agree with the experiment. This equation also predicts that at some particular value of d the denominator will become zero and the current will tend to approach infinity. Actually, of course, as the current gets large, space charges develop which together with the external circuit resistance limit the current to some finite value.

It was stated above that Townsend was able to derive an expression for the observed current increase with electrode separation based on the assumption that the positive ions made ionizing collisions with the gas molecules. It is easy to see now how on the assumption of a mechanism that is admittedly not possible in the majority of discharges the correct current increase could still be predicted. He assumed that each positive ion made β ionizing collisions in advancing unit distance. Then, $\therefore$ the number of positive ions entering a slab of thickness dx at x was P , the number of new ions formed in the slab was

$$d\eta = (\alpha\eta + \beta P) dx \quad (8-21)$$

Using this differential equation and the expression for the total current density

$$J = (\eta + P)q_e \quad (8-22)$$

he derived the following expression for the current:

$$J = J_s \frac{(\alpha - \beta)e^{(\alpha-\beta)d}}{\alpha - \beta e^{(\alpha-\beta)d}} \quad (8-23)$$

This equation exhibits the same properties as Eq. (8-20), and by means of it he calculated values of α and β from experimental data. For the case of

Fig. 8-8 the value of α would be 123.6 as obtained before, and the value of β would be 0.075. This value for β is certainly small in comparison with α , but it represents an ionizing ability for the positive ions very much larger than was ever experimentally observed. It is easy to see now how this paradoxical situation could arise. If we assume that the value of α determined from Eq. (8-23) is slightly different from the value of α determined from Eq. (8-20), we might substitute α' for α in Eq. (8-23). The new equation for J would then be

$$J = J_s \frac{(\alpha' - \beta)e^{(\alpha' - \beta)d}}{\alpha' - \beta e^{(\alpha' - \beta)d}} \quad (8-24)$$

Then, if we make the following substitutions in Eq. (8-24)

$$\left. \begin{aligned} \alpha' &= (1 + \gamma)\alpha \\ \beta &= \gamma\alpha \end{aligned} \right\} \quad (8-25)$$

we obtain Eq. (8-20) exactly.

From this we can see that the difference between α' and α may actually be too small to be observed since γ is so small. Furthermore, since current measurements alone will not indicate the source of the secondary ionization, it is easily possible to confuse the two mechanisms. Although the positive ions may actually be producing secondary-electron emission from the cathode, it appears from Eq. (8-24) as though they are creating ionization by collision in the gas. The value of β obtained, however, is simply $\gamma\alpha$. Although we find experimental values for both γ and β in the scientific literature, Eq. (8-20) is the most commonly accepted relation for the dependence of current on electrode separation, and when values of β appear they are interpreted as being equal to $\gamma\alpha$.

With a particular discharge, it is usually impossible to say exactly what mechanism causes the observed current amplification. In most cases, it is due to a combination of many effects. However, the importance of the causes for the observed current amplification can be easily appreciated if we realize that the solution of this problem determines also the conditions necessary for breakdown or sparking or the transition from a non-self-maintaining to a self-maintaining discharge. A direct consequence of this solution would be a great simplification in the design of mercury-arc rectifiers, high-voltage circuit breakers, lightning arresters, and all types of devices depending for their successful operation on the initiating of a self-maintaining discharge or the prevention of the formation of a self-maintaining discharge except under particular conditions.

What do we mean by a self-maintaining or a non-self-maintaining discharge? Consider again Eq. (8-20) [or Eq. (8-23)] above. For a given value of field strength and gas pressure, α and γ

(or β) are determined and constant. Therefore, as long as d is not too large, J will be directly proportional to J_0 . If we remove the external light source, the current will drop to zero and the discharge will stop. This is a non-self-maintaining discharge, for it depends for its existence on an external ionizing agent. But as the separation of the electrodes increases, the denominator of the fraction approaches zero and the current tends to increase without limit. The value of anode-to-cathode distance, which makes Eq. (8-20) or (8-23) infinite, determines the condition for the establishment of a self-maintaining discharge. At first glance, it would seem to be impossible to have a steady state under these conditions. Actually, however, the current will be limited by the formation of space charges or the capacity of the generator supplying the circuit or some other factor affecting the original conditions which implied a uniform electric field or at least an unvarying electric field and an infinite generator capacity.

We see, then, that the condition for a self-maintaining discharge is given by

$$e^{ad} = \frac{1 + \gamma}{\gamma} \quad (8-26)$$

When this condition is satisfied, the initial source of electrons is no longer necessary and can be removed without disturbing the discharge; i.e., J_0 may also be zero. *Actually the establishment of a self-maintaining discharge means that each electron leaving the cathode is able to produce enough ionization by collision so that the positive ions so formed in returning to the cathode will reproduce one electron.*¹ In this condition the initial source of ionization is no longer necessary, and only then do we have a self-sustaining discharge.

8-6. The Ionizing Coefficients α and γ .—It is worth while now to review briefly the interpretation of α and γ , especially with reference to their application to the Townsend discharge. In

¹ That this condition is implied in Eq. (8-26) is shown as follows: First rewrite Eq. (8-26) in the form $\gamma(e^{ad} - 1) = 1$. We recall that one primary electron will produce e^{ad} electrons in traveling a distance d [Eq. (8-10)]; therefore, one primary electron will produce $(e^{ad} - 1)$ new positive ions. Multiplication of $(e^{ad} - 1)$ by γ gives the number of secondary electrons released by these positive ions. Thus, we see that, when $\gamma(e^{ad} - 1) = 1$, all the positive ions formed by one electron leaving the cathode will be just sufficient to reproduce another electron.

Secs. 8-3 and 6-14, we discussed the dependence of α and γ on the energy of the ions. Ordinarily, in a discharge, we do not know the energy of the ions, but we do know the field strength and the pressure. Because the relationship between the energy of an ion and the field strength is not a simple one but depends in general on the mean free path and because, in a given electric field, all ions will not have the same energy, it is desirable to determine the direct dependence of the average ionizing efficiency on the field strength.

We saw in Sec. 7-1 that the ion does not lose in each collision all the energy that it gains from the electric field. Consequently, its energy continues to increase until the average energy lost in a collision is equal to the average energy gained from the electric field. Its ultimate energy is then given by

$$\bar{W}_m = \frac{\bar{W}_M}{2} + \left[\frac{\bar{W}_M^2}{4} + \frac{q^2 F^2 \lambda_0^2 (m + M)^2}{5.32 m M} \right]^{\frac{1}{2}} \quad (7-10)$$

In the case of electrons the term involving F is much larger than the other terms even with small values of F . Furthermore, in the case of electrons, λ_0 is equal to the mean free path of the electrons. Thus, Eq. (7-10) simplifies to

$$\therefore \bar{W}_m = \frac{q_e F \lambda_e}{2.31} \sqrt{\frac{M}{m}} \quad (8-3)$$

It is obvious, then, that since α , the number of ionizing collisions made by the electron in advancing unit distance, depends on the energy of the electron, it will depend on some function of $(F\lambda)$. It will also be directly proportional to the number of collisions made by the electron in advancing unit distance, *i.e.*, $1/\lambda$. Combining these effects, we get

$$\alpha = \frac{1}{\lambda} f(F\lambda) \quad (8-27)$$

but since λ is inversely proportional to the pressure,¹

$$\alpha = p f\left(\frac{F}{p}\right) \quad (8-28)$$

¹ In all this discussion the temperature is assumed to be constant. Since the mean free path is proportional to the temperature and inversely proportional to the pressure, it is obvious how the effect of temperature variations can be brought into Eq. (8-28) when necessary.

or

$$\frac{\alpha}{p} = f\left(\frac{F}{p}\right) \quad (8-29)$$

That this relation is satisfied was shown by Townsend, when he found that α/p plotted against F/p gave a smooth curve even when the data were taken from widely differing experiments.

On the basis of the kinetic theory, it is possible to derive¹ an expression for the functional equation above of the form

$$\frac{\alpha}{p} = Ae^{-\frac{B}{F/p}} \quad (8-30)$$

For small values of F/p , this equation can be made to fit the experimental data very closely. It is thus a useful expression for computing α under conditions of pressure and field strength where accurate experimental results are not available. As F/p increases, however, the exponential relation ceases to hold true and we must resort to experimental data for the determination of α . There is apparently no single function or group of functions that can express the dependence of α/p on F/p throughout the whole range of experimental possibility.

Figures 8-9 and 8-10 show the experimentally determined variation of α/p with F/p for a few commonly used gases.² In accordance with custom, p is expressed in terms of millimeters of mercury and F in volts per meter. Thus, F/p is a hybrid unit, volts per meter per millimeter mercury. It is evident from these figures that α/p increases quite rapidly with F/p at first; but soon the curve flattens off, and further increases in F/p produce only a slight increase in α/p . Since it has been demonstrated³ that a very slight amount of mercury contamination in the discharge tube can cause a large variation in α/p , these curves may not agree exactly with data taken in highly purified gases.

With Figs. 8-9 and 8-10 before us, it is possible to see why the current obtained in a gas-filled phototube under conditions of constant voltage and constant illumination does not increase

¹ TOWNSEND, J. S., "Electricity in Gases," p. 291, Oxford University Press, New York, 1915.

² A derivation of these curves from data such as those given in Figs. 8-4 and 8-5 is extremely difficult, and as yet there has been no completely successful derivation.

³ General reference 8-7, p. 356.

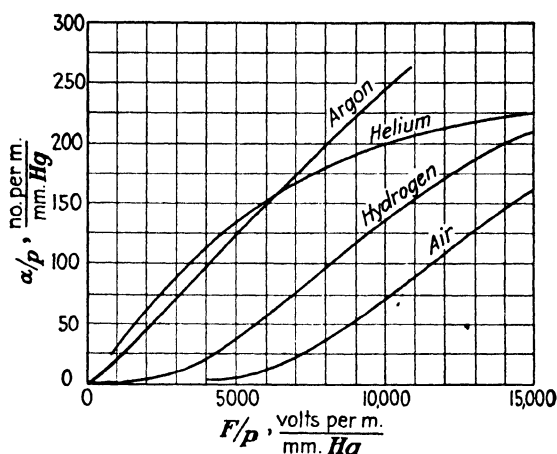


FIG. 8-9.—Ionizing ability of electrons in various gases as a function of field strength and pressure. (Data for air and helium taken from "International Critical Tables," McGraw-Hill Book Company, Inc., New York, Vol. 6, p. 121, Data for hydrogen and argon taken from "Handbuch der Experimentalphysik," Wien-Harms, Leipzig, Vol. 13, Part 3, p. 112, 1929.)

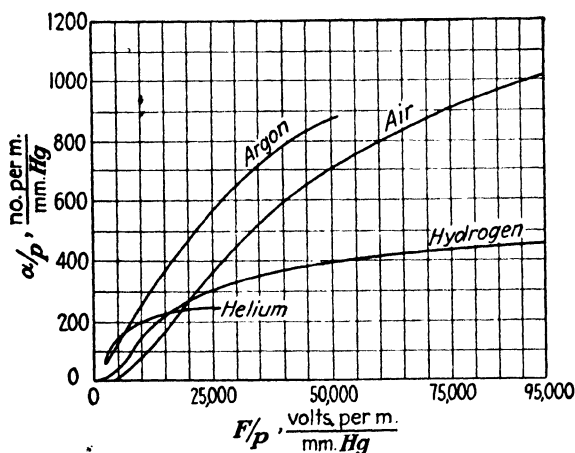


FIG. 8-10.—Ionizing ability of electrons in various gases as a function of field strength and pressure. (Data for air and helium taken from "International Critical Tables," McGraw-Hill Book Company, Inc., New York, Vol. 6, p. 121. Data for hydrogen and argon taken from "Handbuch der Experimentalphysik," Wien-Harms, Leipzig, Vol. 13, Part 3, p. 112, 1929.)

continuously with the gas pressure. As indicated in Fig. 6-45, the current at first increases with an increase in pressure; but when the pressure exceeds a few millimeters of mercury, any further increase in gas pressure results only in a decrease in the current flowing. Since the ionizing coefficient α depends on the pressure, as well as the field strength, it is easy to see how this can occur. Thus, as the pressure increases, the number of collisions made by an electron in traveling unit distance increases because of the decrease in its mean free path, but the energy gained by the electron in a free path decreases. The result is that α varies with the pressure and passes through a maximum when the pressure is a few millimeters.

The conditions necessary to make α a maximum can be determined from the curves of α/p in Figs. 8-9 and 8-10 as follows: Differentiating Eq. (8-28) with respect to p ,

$$\alpha = pf\left(\frac{F}{p}\right) \quad (8-28)$$

Then,

$$\frac{d\alpha}{dp} = f\left(\frac{F}{p}\right) + p\left(-\frac{F}{p^2}\right)f'\left(\frac{F}{p}\right)$$

or

$$\frac{d\alpha}{dp} = f\left(\frac{F}{p}\right) - \frac{F}{p}f'\left(\frac{F}{p}\right) \quad (8-31)$$

But $d\alpha/dp$ is equal to zero when α is a maximum. Therefore, the right side of Eq. (8-31) is also zero when α is a maximum. Thus, since $f(F/p) = \alpha/p$, we have

$$\frac{\alpha}{p} - \frac{F}{p}f'\left(\frac{F}{p}\right) = 0 \quad (8-32)$$

or

$$\frac{\alpha/p}{F/p} = f'\left(\frac{F}{p}\right) \quad (8-33)$$

That is, the value of p for which α is a maximum can be determined from curves similar to those in Figs. 8-9 and 8-10, for at the maximum a tangent to the α/p curve passes through the origin. The method is illustrated in Fig. 8-11. At the point P a tangent to the α/p curve passes through the origin. That is, $\alpha/p = OY$ and $F/p = OX$. Therefore, $f'(F/p)$, or the slope of

the curve at P is

$$f' \left(\frac{F}{p} \right) = \frac{OY}{OX} = \frac{\alpha/p}{F/p}$$

and the conditions for α to be a maximum are determined.

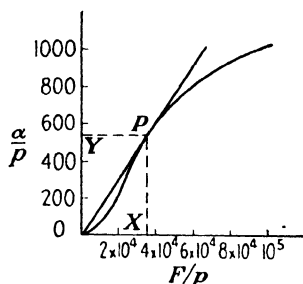


FIG. 8-11.

Since the number of secondary electrons emitted from a surface is, as we saw in Sec. 6-14, a function of both the energy of the incident positive ions and the surface material, we should expect γ to depend on the field strength and the pressure in nearly the same way as α depends on these quantities. There is this difference, however, α depended on the number of collisions made by the electrons in advancing

unit distance and was, consequently, proportional to $1/\lambda$. In the case of secondary emission, γ will depend only on the average

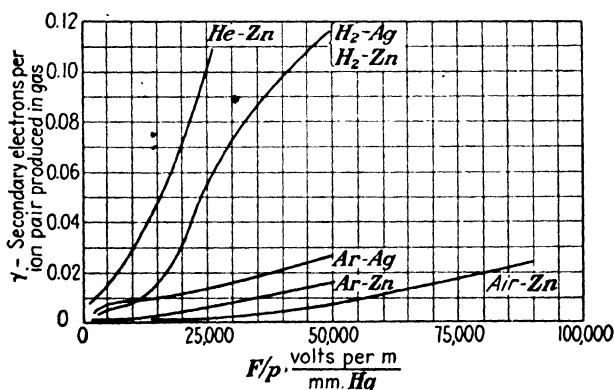


FIG. 8-12.—Secondary electron emission at the cathode as a function of field strength and pressure. (Calculated from data for α/p and β/p in "International Critical Tables," McGraw-Hill Book Company, Inc., New York, Vol. 6, p. 122, and "Handbuch der Experimentalphysik," Wien-Harms, Leipzig, Vol. 13, Part 3, p. 112, 1929.)

energy of the ions produced in the gas and not at all on the number of impacts per meter. Thus, we can write

$$\gamma = g \left(\frac{F}{p} \right) \quad (8-34)$$

Again we must resort to experimental data to determine this functional relation, for the dependence of γ on the cathode material and the field strength is much too complicated to be represented by a single formula. Figure 8-12 gives average values of γ for air, helium, hydrogen, and argon when zinc cathodes were used and for argon and hydrogen when a silver cathode was used. These curves should be interpreted as giving only approximate values for γ . Actually, of course, as can be seen especially in the case of argon, the use of particular cathode materials will vary the results over quite wide limits. The curves are useful, however, for making order-of-magnitude calculations.

Problems

1. (a) What percentage of mercury atoms would be thermally ionized if the gas temperature were raised to 500°C .? (b) Recalculate for 1000°C . (c) Recalculate for 3000°C . Assume atmospheric pressure.
2. Calculate the temperature equivalent of the average energy of electrons in mercury vapor at a pressure of 0.1 mm. Hg and a temperature of 90°C . when the field strength is 10 volts/m. Assume that the mean free path of mercury is 14×10^{-8} m. at 20°C . and 760 mm. Hg.
3. What proportion of the electrons in Prob. 2 will have energies large enough to ionize mercury atoms? HINT: Assume a Maxwellian distribution of electron speeds.
4. How many ionizing collisions per meter will an electron make in moving through a discharge in argon when the pressure is 1 mm. Hg, the temperature is 0°C ., and the electrons have an energy of 25 volts?
5. Assuming that X rays uniformly ionize the space between a pair of parallel plates separated a distance of d m., derive the expression for the current density when ionization by collisions of electrons with gas atoms occurs [Eq. (8-12), Sec. 8-5].
6. Assume that a photoelectric tube contains helium at a pressure of 0.5 mm. Hg. Further assume that the electrodes are parallel plates and that light falling on the cathode makes available a maximum electron current density of 0.5×10^{-2} amp./sq. m. If the intensity is held constant at 5,000 volts/m., find how the current density through the tube varies as the distance between the electrodes varies from 0.5 to 3 cm. How must the voltage change in order to maintain the required intensity? Plot curves of voltage and current density as a function of electrode separation.
7. A phototube similar to that in Prob. 6 contains argon at a pressure of 0.5 mm. Hg. If the gap between the electrodes is 1 cm. and the saturation emission of the cathode is 3×10^{-2} amp./sq. m., calculate the current density through the tube at impressed voltages of (a) 30 volts, (b) 75 volts.
8. Assume that the space between two parallel electrodes in hydrogen is uniformly ionized by X rays. What will be the value of the current density in the tube if the pressure is 0.25 mm. Hg, the gap is 2 cm., the potential applied is 75 volts, and the saturation current is 10^{-2} amp./sq. m.?

9. A parallel-plate phototube with silver cathode contains argon at a pressure of 0.5 mm. Hg. If the electrode spacing is 1 cm. and the saturation emission from the cathode is 10^{-2} amp./sq. m., how does the current density vary as the voltage is raised from 25 to 225 volts, on the assumption that only ionization by collisions of electrons with gas atoms occurs? Sketch a curve.

10. Assuming the same conditions as in Prob. 9, how would the current density vary with the voltage if secondary emission at the cathode due to positive-ion bombardment also occurred?

11. Two parallel zinc plates in a hydrogen-discharge tube are separated a distance of 2 cm. If the pressure in the tube is $\frac{1}{2}$ mm. Hg and the saturation current density due to photoelectric emission is 10^{-2} amp./sq. m., how will the current density vary with the voltage in the range between 50 and 450 volts? At what potential is a self-maintaining discharge formed? Sketch the current-voltage curve.

12. Show that when the space between two plane-parallel electrodes is uniformly ionized by X rays or cosmic rays the current is given by the following equation:

$$J = J_s \left[\frac{1 + \gamma}{\alpha d} \right] \left[\frac{e^{\alpha d} - 1}{1 - \gamma(e^{\alpha d} - 1)} \right]$$

13. Given two plane-parallel zinc electrodes in air at a pressure of 2 mm. Hg and with a separation of 0.5 cm. It being assumed that the saturation current density due to cosmic rays and natural radioactivity is 10^{-8} amp./sq. m., how does the current density in the gas vary with the impressed voltage in the range from 25 to 425 volts, (a) when electrons alone produce ionization, and (b) when electrons and positive ions produce ionization? (c) Plot a curve of $\log J$ vs. potential. (d) At what potential is a self-maintaining discharge formed?

General References

- 8-1. THOMSON, J. J., and G. P. THOMSON: "Conduction of Electricity through Gases," 3d ed., Vol. I, pp. 446-473, Vol. II, Chaps. III and IV, also pp. 222-277, 512-523, Cambridge University Press, London, 1928.
- 8-2. DARROW, K. K.: "Electrical Phenomena in Gases," pp. 3-18, 21-23, 35-44, 47-59, 71-72, 271-283, The Williams & Wilkins Company, Baltimore, 1932.
- 8-3. COMPTON, K. T., and I. LANGMUIR: "Electrical Discharge in Gases," *Rev. Mod. Phys.*, **2**, 124-131, 134-136, 1930.
- 8-4. v. ENGEL, A., and M. STEENBECK: "Elektrische Gasentladungen," Vol. I, pp. 29-108, 119, Verlag Julius Springer, Berlin, 1932.
- 8-5. LOEB, L. B.: "Kinetic Theory of Gases," 2d ed., pp. 623-635, McGraw-Hill Book Company, Inc., New York, 1934.
- 8-6. LOEB, L. B.: "Mechanism of the Static Spark Discharge," *Rev. Mod. Phys.*, **8**, 267ff., 1936.
- 8-7. LOEB, L. B.: "Fundamental Processes of Electrical Discharge in Gases," Chaps. VIII and IX, John Wiley & Sons, Inc., New York, 1939.

CHAPTER IX

SPARKING POTENTIAL, SPARK-OVER, AND CORONA DISCHARGE¹

9-1. Types of Discharge.—In many situations found in engineering practice, it is desirable to know how much voltage can be impressed on a pair of electrodes separated by a gas space before a discharge will occur. Such widely different examples can be cited as the gap in a low-pressure gaseous glow-discharge tube, and the gap in air between the wires of a high-voltage transmission line. In Sec. 8-5 the theories relating to the passage of a current between parallel plates were developed. In the present chapter, the theory will be generalized to apply to the situations found in practice. The experimentally determined relations applying to sparking potential and corona discharge will also be examined.

When the voltage on a pair of electrodes immersed in a gas is raised to a high enough value, a luminous discharge occurs. In air at atmospheric pressure, this discharge often takes the form of a bright flash, and the flash is accompanied by a sharp crackling or explosive sound. This constitutes the well-known spark-over, or breakdown. With a large source of power and a low electrical circuit resistance the discharge following the spark will consist of an arc. The arc discharge is characterized by a relatively low voltage between the electrodes. The changes described above can be illustrated by an idealized current-voltage curve as shown in Fig. 9-1. This figure illustrates the fact that preceding the spark-over there is a small dark discharge current. This current is caused by ionization by collision in exactly the same way as that described in Sec. 8-5. The only difference lies in the assumption regarding the initial supply of electrons or ions required to start the process. In ordinary air the initial electrons and ions are continuously supplied by ionization due to cosmic rays and radioactive radiation. If any ultraviolet light strikes

¹ General references for Chap. IX will be found on page 310.

the cathode, there will also be a supply of electrons released by the photoelectric effect. When the voltage on the gap reaches a certain value, shown by the ordinate E_s in Fig. 9-1, the ionizing processes become unstable, the current concentrates into a small

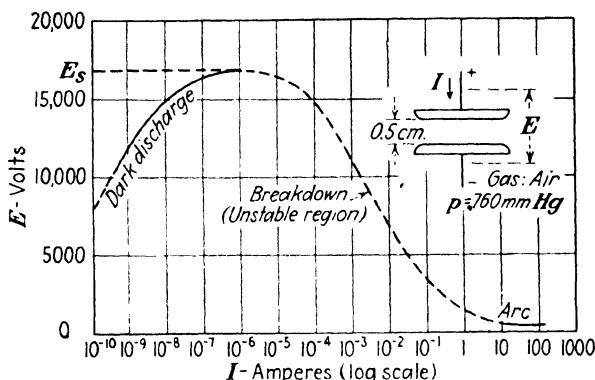


FIG. 9-1.—Schematic current-voltage curve.

path, and a spark passes across the gap. The potential difference E_s at which the discharge becomes unstable and suddenly changes its form will be called the *sparking potential*.

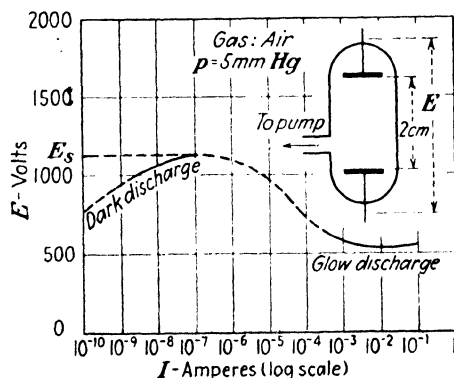


FIG. 9-2.—Current-voltage curve illustrating the sparking potential E_s .

Although the behavior shown in Fig. 9-1 is typical of many discharges, there are other important discharges in which the dark discharge is not followed immediately by an arc, but rather by a glow discharge, or a corona, or brush, discharge. A typical case in which a glow discharge is obtained is illustrated in Fig. 9-2. The voltage E_s at which the transition to a glow takes place is again

called the *sparking potential*. The transition to the glow or arc is accompanied by much stronger ionization and consequent changes in the electric field caused by space charges. The currents in the dark discharge are so small that there is little or no distortion of the field due to space charges. The main difference between the dark and the glow, or arc, discharges is that these discharges are self-maintaining whereas the dark discharge is non-self-maintaining. Thus, the glow discharge would be maintained even though the source of initial electrons and ions were removed, whereas the dark-discharge current would drop to zero if no initial electrons or ions were formed in the gap.

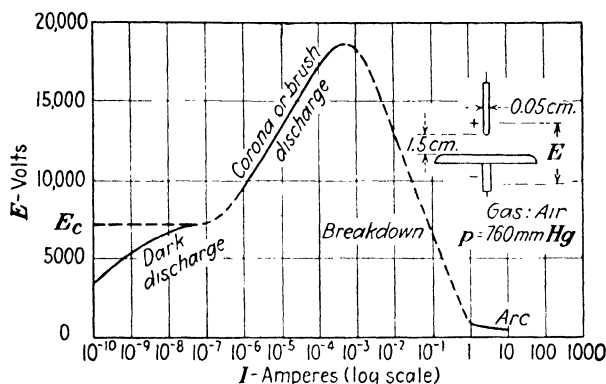


FIG. 9-3.—Current-voltage curve illustrating the corona voltage E_c .

The glow and arc discharges will be discussed in detail in Chaps. X and XI.

In cases where one electrode has a small radius of curvature compared with the gap length the dark discharge is succeeded by a discharge in which the voltage increases with increased current, as shown in Fig. 9-3. This figure illustrates the formation of a corona discharge on a thin rod ending in a hemispherical surface. The discharge assumes a variety of shapes, depending on the pressure of the gas, the polarity of the electrode, and the impressed voltage. The discharge is usually accompanied by a hissing sound. The light given off is characteristic of the gas and for air has a violet or reddish-purple color. The change from the dark discharge to the visible discharge may or may not be accompanied by an unstable transition in the current-voltage curve. For example, when the point electrode in Fig. 9-3 is positive, there is a fairly smooth transition in the current curve;

but when the point has a negative polarity, there is an unstable transition region. For practical purposes, it suffices to define the voltage E_c at which a visible glow appears as the *corona voltage*. This definition can be made somewhat more precise by specifying that the observations should be made in perfect darkness by an observer whose eyes have become adjusted to maximum sensitivity.

The current density in the corona discharge is sufficient to alter the field distribution because of the space charges present. As the voltage increases and the current increases also, a point is reached at which strongly ionized *streamers* form in the discharge. In long gaps at atmospheric pressures the longer streamers take on a treelike or branching form. These are called *brush discharges*.¹

When the voltage is raised still more, an ionized path finally bridges the gap and complete breakdown results. The maximum voltage impressed on the gap before breakdown results is called the *breakdown voltage*.

The dark-discharge current in the instances discussed above may be considered to be somewhat variable with time, for the supply of initial electrons and ions is not a constant. For this reason, it might be expected that if the potential on a gap is rapidly increased the breakdown potential may be different from that obtained with a very slowly increasing potential. This is actually found to be true, and a higher breakdown potential is obtained with a given gap with a rapidly increasing voltage than with a slowly rising voltage. The magnitude of this effect depends on the sources of initial ionization present, on the pressure of the gas, and on the rate of rise of the potential. The sparking and breakdown potentials as defined in the preceding paragraphs will refer to the value obtained with a slowly increasing potential.

9-2. Sparking Potential in a Uniform Field. Paschen's Law. The physical processes that *can* lead to a self-maintaining discharge and hence to sparking in a uniform field can be clearly

¹The definitions and nomenclature relating to discharges have not yet been standardized; for example, the term *sparking potential* is sometimes applied only to those cases in which a complete breakdown to an arc occurs. According to our more general definition of sparking potential, we also include the transition to a self-maintaining condition in the corona as occurring at the sparking potential.

outlined, but the detailed processes actually active in particular situations are not always definitely known. The extensive investigations¹ of recent years have tested the older hypotheses and have gradually led to a more comprehensive, though somewhat more complicated, theory of sparking.

In this theory the process of ionization by collision by electrons plays a most prominent part; but this process cannot, of itself, lead to a spark. As explained in Sec. 8-5, secondary processes can generate new ions which amplify the ionization by collision and thus lead to a self-maintaining discharge and to a spark. The secondary processes that can release electrons at the cathode are the following: (1) positive-ion impact, (2) photoelectric effect, and (3) impact of metastable atoms. The secondary processes that might be active in the gas are listed as follows: (4) photoelectric effect, (5) field distortion by space charges, (6) thermal ionization, and (7) ionization by metastable atoms.

Under any particular experimental condition, one or more of these secondary processes may be active, and the question of their relative importance arises. Although this question has been the subject of considerable controversy, some general conclusions are possible. Thus, in short gaps and at low pressures the secondary processes at the electrodes are undoubtedly more important than those in the gas space. Of the processes at the cathode the impact of metastable atoms is effective only in the rare gases and in mercury vapor, so that, in most common gases, it is the positive-ion impact or the photoelectric effect that supplies the secondary electrons. In longer gaps at pressures above, say one-tenth atmospheric, the secondary processes in the gas itself may lead to a self-maintaining discharge. Recent investigations of Loeb and his coworkers² show that secondary processes 4 and 5, *i.e.*, photoionization in the gas and field distortion caused by space charges, are probably important.

The foregoing discussion leads to the conclusion that, at least in the common gases at low pressures and in short gaps, the conditions preceding sparking are similar to those discussed in Sec. 8-5. Consequently, the criterion for a self-maintaining discharge and therefore for the sparking potential is given by Eq. (8-26) (Sec. 8-5), *i.e.*,

¹ See general references 9-5, 9-6, 9-7, and 9-8.

² General references 9-5, 9-6, and 9-7.

$$e^{\alpha d} = \frac{1 + \gamma}{\gamma} \quad (8-26)$$

where d is the gap length, α is the ionizing coefficient for electron impacts in the gas, and γ is the ionizing coefficient for positive-ion or photon impacts at the cathode. Equation (8-26) must be regarded as an approximate criterion for sparking, for two reasons. In the first place, it should be recalled that the expression is based on a steady initial ionization at the cathode corresponding to the current J_s in Eq. (8-20), whereas actually the natural primary ionization is extremely variable. This objection is not serious if slowly varying voltages only are involved, for the initial ionization is then sufficient to provide the primary electrons near the cathode. Secondly, it should be noted that the current at the sparking potential cannot become infinite as implied in the derivation of Eq. (8-26). However, the current is changing so rapidly with voltage near the sparking potential that the value of E_s corresponding to the actual finite current would be only slightly below the value of E_s predicted by Eq. (8-26).

In general, γ is small compared with unity so that Eq. (8-26) reduces to the approximate form

$$e^{\alpha d} = \frac{1}{\gamma} \quad (9-1)$$

This equation could be used to calculate the sparking potential if the functions for α and γ given in Eqs. (8-28) and (8-34) were known. Since the functions $f(F/p)$ and $g(F/p)$ are not usually known in a mathematical form, a theoretical equation showing how the sparking potential depends on the gap length and the pressure is seldom derived. Nevertheless, a very useful functional relation among these variables can be found by substituting for α and γ in Eq. (9-1) their values given in Eqs. (8-28) and (8-34). After making this substitution and after replacing the intensity by its equivalent¹ E_s/d , where E_s is the sparking potential, we find

$$e^{pd f\left(\frac{E_s}{pd}\right)} = \frac{1}{g(E_s/pd)} \quad (9-2)$$

In this equation the pressure and gap length appear only as their product pd . If the mathematical forms of the functions f and g

¹ The intensity in the uniform field between parallel plates is the voltage divided by the gap length.

were known, it would theoretically be possible to solve for the relation between E_s and the pd product. Therefore, Eq. (9-2) shows that the sparking potential is a function only of pd , thus:

$$E_s = f'(pd) \quad (9-3)$$

This equation is the mathematical statement of Paschen's law, which states that the *sparking potential in a uniform field is a function only of the product of gas pressure and the gap length.*

Paschen's law, as stated above, holds true only as long as the temperature remains constant. The effect of the temperature of the gas can readily be included if we recall the derivation of the functional relations for α and γ . The fundamental quantities that determine α and γ are the intensity and the mean free path λ . Equations (8-28) and (8-34) were derived by replacing λ by $1/p$; if the effect of the temperature is included [see Eq. (4-53)], λ is replaced, except for a proportionality constant, by T/p . But T/p is proportional to $1/\delta$, where δ is the density of the gas. Thus, the density δ takes the place of the pressure p , and the revised form of Eq. (9-3) is therefore,

$$E_s = f''(\delta d) \quad (9-4)$$

In this form, Paschen's law states that *in a uniform field the sparking potential is a function only of the product of the gas density¹ and the gap length.* This can be interpreted to mean that for a given gas the sparking potential is a function only of the *mass* of gas between the electrodes.

Paschen's law is particularly useful for correlating the results of experiments on sparking potentials. The law has been verified by extensive tests, covering pressures from about 0.5 mm. Hg to about 10 atm., and gap lengths from 1 mm. to several centimeters. The curves in Fig. 9-4 show the average of the results obtained by several investigators for the sparking potential as a function of pd for air, hydrogen, and carbon dioxide. The more general form of Paschen's law given in Eq. (9-4) agreed with

¹ The density is often measured relative to the value $\delta = 1$ at 760 mm. Hg and 0°C. As long as the gas follows the perfect-gas laws, the *relative* density will be given by

$$\delta = \frac{273p}{T}$$

where p is in atmospheres and T in degrees Kelvin.

the measurements of Bowker¹ for dry hydrogen and nitrogen up to 860°C.

A notable feature of the curves in Fig. 9-4a is the minimum value in the curve and the rising value of voltage for smaller values of pd . The potential at the minimum is called the *minimum sparking potential*, and the corresponding value of the

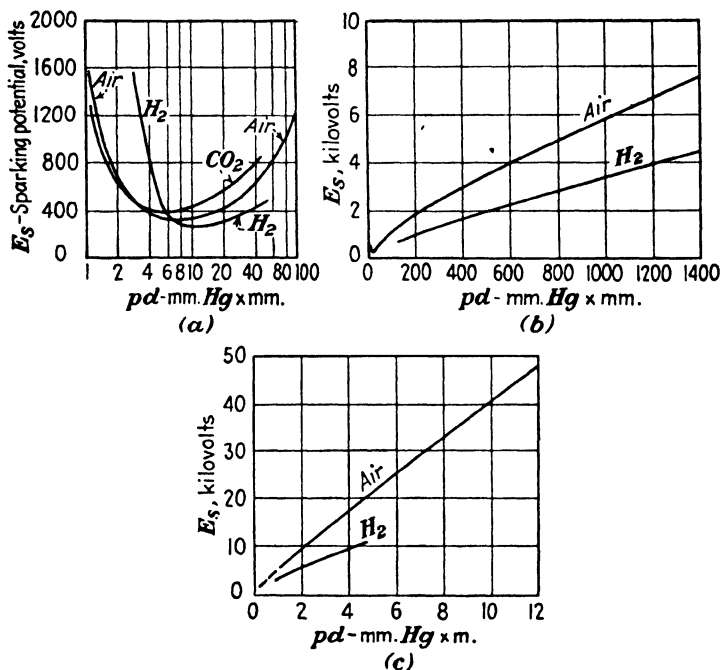


FIG. 9-4.—Sparkling potentials between parallel-plate electrodes as a function of pd , at a temperature of 20°C. (General reference 9-3, pp. 53-57.)

abscissas is called the *critical pd product*. A qualitative explanation of the minimum in the curve is found in the small number of ionizing collisions an electron or ion makes in crossing the gap. In the neighborhood of the critical pd , the mean free path is of the same order of magnitude as the gap length, and an electron may make only 10 to 20 collisions in crossing the gap. Of these collisions, relatively few will be ionizing collisions. Consequently, as the pressure is reduced still more (a constant gap length being assumed), the voltage on the gap must be increased

¹ *Proc. Phys. Soc. London*, **43**, 96, 1931.

to increase the percentage of the collisions that result in ionization.

When the pressure is increased so that pd is larger than the critical value, the sparking potential rises on a nearly linear curve. In this range the number of collisions increases, but the mean free path of the electrons decreases with an increase in the pressure, so that the intensity, and hence the voltage, must increase in order to produce sufficient ionization by collision to establish a self-maintaining discharge.

The theoretical criterion for sparking [Eq. (9-1)], contains the factor γ which depends on the metal of the cathode. It would be expected, then, that the sparking potential would depend on the electrode metal. In the neighborhood of the minimum sparking potential, this effect has been observed, for example, by Ehrenkrantz,¹ who found a 50 per cent reduction in the minimum sparking potential for argon upon allowing sodium to deposit on the platinum cathode. However, the change in the sparking potential at a value of pd of 240 mm. Hg $\times$ mm. was only 7.5 per cent. When various common metals are used for the cathode, the effect on the sparking potential is much smaller and is negligible near atmospheric pressure for gaps longer than 2 or 3 mm. This can be explained, at least for the lower pressures, by the slow variation of the sparking potential with the value of γ as calculated by Eq. (9-1). A numerical calculation will be helpful to show how small an effect a change in γ would have on the sparking potential according to the theory. Consider a 10-mm. gap in air at a pressure of 10 mm. Hg. Then, in Fig. 9-4a, we find the experimental value of E_s at $pd = 100$ to be 1,230 volts. Now, evaluate the left side of Eq. (9-1). To find α , we calculate F/p to be 12,300 and then find α/p from Fig. 8-9, finally obtaining a value of 1,150 for α . Thus, $e^{\alpha d}$ has the value 100,000; and if Eq. (9-1) is satisfied, γ must have the value $1/100,000$ or 1.0×10^{-5} . This value for γ is too low compared with the values obtained experimentally as given in Fig. 8-12, but it will suffice for the present argument. Now, suppose that a change of cathode metal changes γ by a factor of 2, i.e., to 2×10^{-5} . The corresponding α needed to satisfy Eq. (9-1) is now 1,082, a value giving only a 5.9 per cent change in α . The percentage change in the sparking potential is less than 5.9 per cent, for according to Fig. 8-9 the given change in

¹ *Phys. Rev.*, **55**, 219, 1939.

α/p is accompanied by a smaller percentage change in F/p and hence by a smaller change in E_s .

• Paschen's law was deduced from Eq. (9-1), which in turn is based on the secondary process of electron emission at the cathode by the impact of positive ions and photons, or the " γ -mechanism" as it may be called. The fact that Paschen's law holds for pressures of the order of 1 atm. and higher and for gap lengths of the order of 1 cm. should not, however, be construed to imply that the γ -mechanism is necessarily the only one acting under these conditions. One objection to the γ -mechanism is that it cannot successfully account for the extremely short breakdown times of the order of 5×10^{-8} sec. observed in time-lag studies of breakdown. These breakdown times are so short that the positive ions would not have time to reach the cathode to liberate secondary electrons. The liberation of electrons at the cathode by photons generated in the gas is likewise too slow a process to explain the very short-time breakdowns, although this process would be many hundred times as fast as that involving positive-ion impact. The secondary processes that *can* account for the short-time breakdowns are photoionization in the gas caused by photons from the ionized region and, secondly, distortions of the field by space charges that promote subsequent ionization. These processes, operating either singly or in conjunction,¹ are considered to account for sparking at pressures above, say 100 mm. Hg, and for gap lengths over 2 or 3 mm. A brief account of the mechanism for the initiation of self-maintaining discharges involving these processes will be presented in Sec. 9-6.

9-3. Sparking Intensity in a Uniform Field.—The electric intensity in the uniform field that corresponds to the sparking potential is called the *sparking intensity*.² The experimental curves in Fig. 9-4 can readily be interpreted in terms of the electric intensity in the gap. For example, consider the variation of the sparking intensity of air with the pressure. If a constant gap length is assumed, then the ordinates of the curves in Fig. 9-4 divided by the gap length give the sparking intensity. Therefore, the sparking intensity for a constant gap length varies

¹ The theory of sparking under these conditions is not yet complete enough to provide an exact description of the active mechanisms.

² Also sometimes called sparking gradient, disruptive gradient, or dielectric strength.

with the pressure in the same fashion as the sparking potential varies with pd .

In order to establish the variation of the sparking intensity with the gap length, all that is necessary is to assume a fixed pressure and compute the intensities from the curves in Fig. 9-4. If the curve shown in Fig. 9-4*b* were a straight line through the origin, then the sparking intensity would be independent of the gap length. However, there is a double curvature in the relation shown in Fig. 9-4*b*, and consequently the sparking intensity varies with the gap length. To illustrate this variation, consider

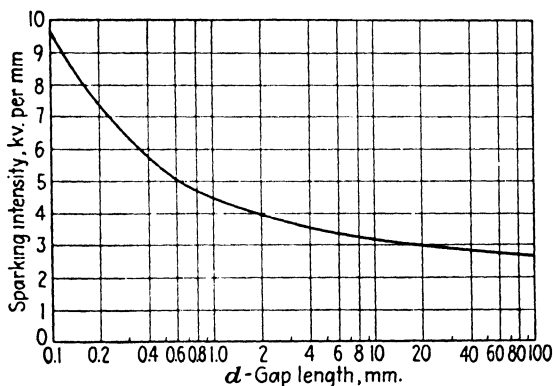


FIG. 9-5.—Sparking intensity in air in a uniform field. $p = 760$ mm. Hg, $T_c = 20^\circ\text{C}$.

the experimental results for an air gap at atmospheric pressure shown in Fig. 9-5. The curve in Fig. 9-5 was drawn by Schumann¹ through a large number of test points obtained by different observers. The curve shows that it is impossible to specify some definite sparking intensity in air, such as the value of 3 kv./mm., which has often been quoted.

The descending nature of the curve in Fig. 9-5 could have been predicted by making use of Eq. (9-1). This can perhaps be done most simply by assuming, for the moment, that the sparking intensity remains constant as the gap length is increased. Since F has been assumed constant, the values of α and γ would remain constant and it is obviously impossible to maintain the equality of the two sides of Eq. (9-1) as the gap length increases. Therefore, the sparking intensity cannot remain constant but must decrease so that α and γ decrease as d increases.

¹ General reference 9-3, p. 25.

A consideration of the ionizing processes also leads to an explanation of the trend of the curve in Fig. 9-5. The self-maintained discharge is reached when a single electron leaving the region of the cathode causes enough ionizing collisions so that the positive ions (or photons) so formed generate at least one electron in returning to the cathode. As the gap is increased, a single electron leaving the cathode would give rise to more and more ionizing collisions if α remains constant. There would then be more positive ions (or photons) than needed to generate one electron at the cathode to replace each initial electron; consequently α must decrease as the gap increases, and this corresponds to the decrease in sparking intensity.

9-4. Sparking in Vacuum, in Gases under High Pressure, and in Mixed Gases.—Paschen's law ceases to apply at very low and at high pressures. At pressures below 10^{-5} to 10^{-4} mm. Hg the breakdown no longer depends on the kind of gas between the electrodes and its pressure but does depend to a considerable extent on the condition of the electrode surfaces and on the material of the electrode. Moreover, the results of breakdown experiments are erratic on account of the effect of the long-time lags encountered. The order of magnitude of the breakdown voltage in a high vacuum with electrodes not degassed but specially treated to increase the breakdown voltage was found by Anderson¹ to be 120 kv. for stainless steel and 37 kv. for copper for a gap of 1 mm. The treatment of the electrodes consisted of running a hydrogen glow discharge between them and then passing preliminary sparks in a high vacuum. This procedure resulted in a breakdown voltage more than double the initial value.

The initiation of the breakdown in high vacuum might be attributed solely to the emission of electrons by the high-field effect; but this cannot account for all breakdowns, for most electrodes have adsorbed gas layers or oxidized surfaces which facilitate the breakdown.

Considerable interest has been shown in the problem of the breakdown of gases at high pressures, both from the scientific and the practical viewpoint. Compressed-gas insulation has been utilized in high-tension transformers, high-voltage condensers, and in the Van der Graaf high-voltage generator (see

¹ *Elec. Eng.*, **54**, 1315, 1935.

Sec. 3-9). On the scientific side the interest has been partly in a determination of the range of validity of Paschen's law. The investigations¹ showed that above a pressure of about 10 atm. the sparking potential in a uniform field was no longer simply a function of pd . Therefore, several curves must be employed to show the results of tests; for example, the sparking potential is plotted as a function of the pressure with the gap length as a parameter. Sample curves are shown in Figs. 9-6 and 9-7. Howell reported that the sparking potential at the higher pres-

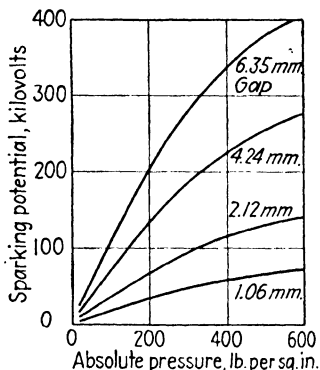


FIG. 9-6.—Breakdown in compressed air in a nearly uniform field, tested with a direct voltage. (Data of A. H. Howell, *Trans. A.I.E.E.*, **58**, 198, 1939.)

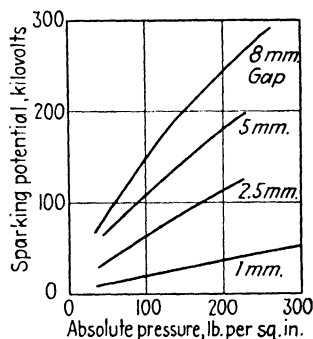


FIG. 9-7.—Breakdown in compressed carbon dioxide. The sparking potential is the peak value of a 50 cycle/sec. wave. (Data of A. Palm, *Arch. Elek.*, **28**, 298, 1934.)

ures is reduced to a value one-half to one-third of the values in Fig. 9-6 by slightly roughening the electrodes. He found that after passing a large number of sparks between the electrodes the sparking voltage returned to its high value. The values in Fig. 9-6 refer to tests using electrodes conditioned by passing sparks until a substantially constant sparking potential was attained.

The effect of metastable atoms on the sparking potential of mixtures of the inert gases has been shown by the experiments of Penning. A very small percentage of admixed argon has a decided effect on the sparking potential of neon as shown by the curves² in Fig. 9-8. The lowering of the sparking potential is caused by the ionization of the argon atoms (ionization energy

¹ See bibliography by A. H. Howell, *Trans. A.I.E.E.*, **58**, 194, 1939.

² PENNING, F. M., and C. C. J. ADDINK, *Physica*, **1**, 1007, 1934.

15.69 e.v.) by the metastable neon atoms of 16.54 e.v. energy. In addition to this action in the gas space, electrons can be released at the cathode by the impact of the metastable neon atoms and by positive ions. Thus, three secondary processes are probably acting simultaneously to build up the ionization. A similar lowering of the sparking potential in He, Ne, and A is caused by a slight admixture of mercury vapor, and this is explained in a similar way. The low sparking potential in mixtures of this sort is helpful in the satisfactory starting of the discharge in fluorescent-tube lamps.

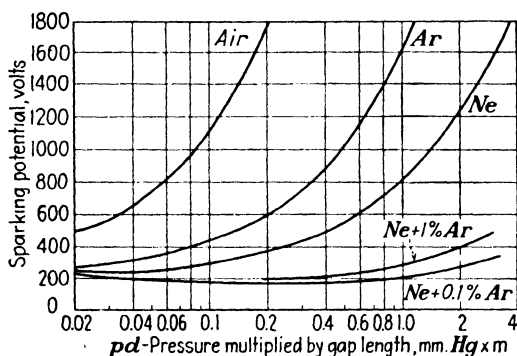


FIG. 9-8.—Sparking potentials for air, argon, neon, and neon-argon mixtures.

With the application to the insulation of high-voltage apparatus in mind, attempts have been made to find gas or gas and vapor mixtures that have high dielectric strengths. For example, Charleton and Cooper¹ tested 77 different gases and mixtures using a gap between nickel electrodes shaped so as to produce a nearly uniform field. Some of their results are reproduced in Table 9-1. The $N_4 + CCl_4$ mixture exhibits a dielectric strength 65 per cent higher than that of N_2 when the total pressure is 1 atm. When the total pressure is 13.5 atm. with a partial pressure of CCl_4 of 0.138 atm. the dielectric strength is 16.2 times that of N_2 at 1 atm. However, when this mixture is compared with pure N_2 at 13.5 atm., it is found that the sparking potential of the mixture is only 19 per cent above that of the pure N_2 gas.

The increase in the sparking potential resulting from the admixture of small amounts of vapor containing the halogens,

¹ *Gen. Elec. Rev.* 40, 438, 1937.

TABLE 9-1.—RELATIVE DIELECTRIC STRENGTHS OF VARIOUS GASES AND NITROGEN PLUS VAPOR MIXTURES
(Temperature = 23°C.)

Gas or mixture	Total pressure, atm.	Partial pressure of vapor, atm.	Relative dielectric strength*
$N_2 + CCl_4$	1	0.138	1.65
$N_2 + CCl_4$	6.05	0.138	7.9
$N_2 + CCl_4$	13.5	0.138	16.2
$N_2 + CHCl_3$	1	0.237	1.58
CCl_2F_2	1		2.4
CCl_2F_2	6.05		15.3
SO_2	1		1.9
CH_2Cl	1		1.06
$N_2 + C_6H_6$	1	0.112	0.84
$N_2 + C_6H_5Cl$	1	0.0138	0.81

* Sparking potential relative to that for N_2 at 1 atm. in the same gap. Gap = $\frac{1}{2}$ to $\frac{1}{4}$ in.

or of O_2 or SO_2 is ascribed to the removal of free electrons from the ionized regions in the dark discharge by electron attachment. The negative ions resulting from the attachment process are no more effective than the positive ions as ionizing agents in the gas space, and so the total ionizing activity is reduced.

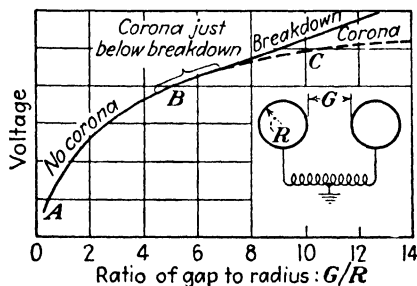


FIG. 9-9.—Breakdown and corona in sphere gaps.

9-5. Sparking, Corona, and Breakdown in Nonuniform Fields.

Nonuniform fields in air at atmospheric pressure are encountered in the sphere-gap voltmeter, between the parallel wires of high-voltage transmission lines, and in many other situations. A large number of investigations have been carried out to find the relation of the sparking potential to the electrode dimensions, the gap length, and the air pressure and temperature for some of these situations. The electrode arrangements that have

received most attention are (1) adjacent spheres, (2) parallel wires, and (3) concentric cylinders (a wire in a cylinder). For small spacings of the electrodes compared with their radius of curvature, a complete breakdown of the gap occurs. For relatively large gap lengths, a partial failure of the gap occurs, resulting in a corona discharge in the most highly stressed parts of the gap. These results¹ are illustrated schematically for adjacent spheres with symmetrically applied voltage in Fig. 9-9. For a given ratio of gap to sphere radius (G/R), if the voltage is

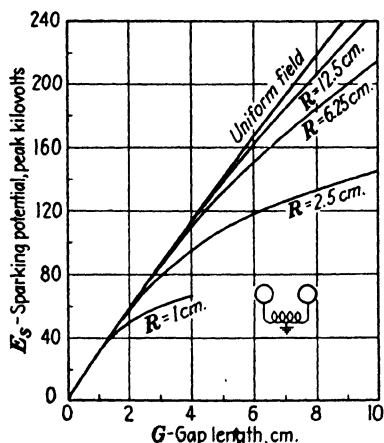


FIG. 9-10.—Sparking potentials for sphere gaps in air at 760 mm. Hg and 20°C. R = sphere radius. (S. Franck, "Messentladungsstrecken," pp. 173ff., Verlag Julius Springer, Berlin, 1931.)

below the curve ABC , then the ionization in the gap must be maintained by an external ionizing agent, although ionization by collision may be present. At the ordinate corresponding to the curve ABC , the discharge becomes self-maintaining, and a corona discharge or a complete breakdown results. Figure 9-9 shows the ranges over which a discharge precedes complete breakdown of the gap.

The variation of the sparking potential for equal, adjacent spheres as a function of the gap length is shown in Fig. 9-10 for various diameters of the spheres.² The curve for parallel plates is shown for comparison. The sparking potentials in the nonuniform fields are lower than in the uniform field.³ Indeed, as the field becomes more nonuniform, *i.e.*, with a given gap length and smaller spheres, the sparking potential decreases. When the maximum electric intensity in the gap corresponding to the sparking potential is computed, the results are found to be

¹ General reference 9-2, p. 68.

² For information on the use of the sphere spark gap for measuring high potentials see A.I.E.E. Standard No. 4, *Measurement of Test Voltage in Dielectric Tests*, 1940.

³ As we shall see presently, however, under low-pressure conditions the sparking potential may be higher in the nonuniform field than in a uniform field.

dependent both on the gap length and on the size of the spheres as shown in Fig. 9-11. The smaller spheres require a higher intensity at their surface to produce sparking than the larger

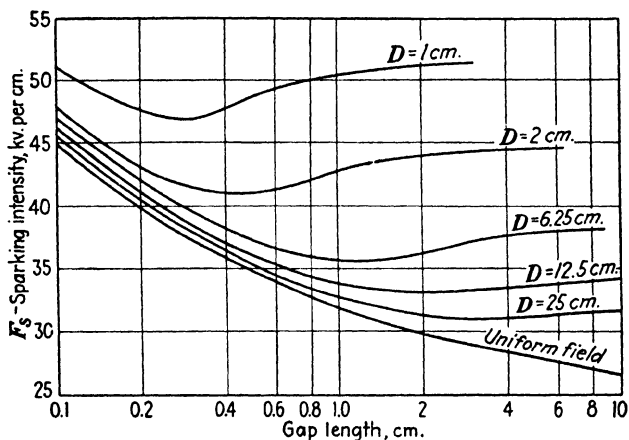


FIG. 9-11.—Sparking intensity as a function of gap length for sphere gaps with symmetrical potentials with respect to ground. D = diameter of spheres. (S. Franck, *Arch. Elek.*, **23**, 226, 1929.)

spheres or parallel plates. This means that in order to satisfy the condition for a self-maintaining discharge the ionization by collision must be enough greater in the strong parts of the field to make up for the deficiency in the weaker parts of the field. The variation of the electric intensity in the gap between the spheres is illustrated in Fig. 9-12 and is compared with the sparking intensity that would be necessary to cause breakdown in a uniform field having the same gap length. This illustrates that the ionization preceding spark-over will be more intense than for the uniform field for a region close to the spheres, regions AB and CD , but will be less intense over the range BC in the figure. If larger spheres are considered, the curve of intensity will have a more gradual slope. Therefore, the region

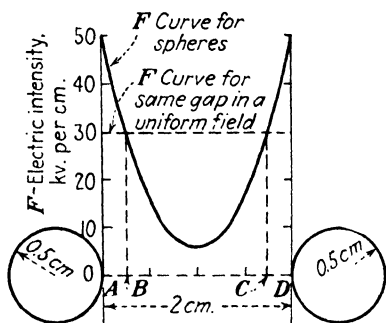


FIG. 9-12.—Intensity curve for a sphere gap compared with the breakdown intensity for the same gap in a uniform field.

of ionization that is more intense than for the uniform field will extend out farther from the spheres, and the maximum intensity does not need to be so high to produce enough ionization to satisfy the condition for a self-maintained discharge.

Under low-pressure conditions the effect of the nonuniform field may be to raise the sparking potential. This occurs in

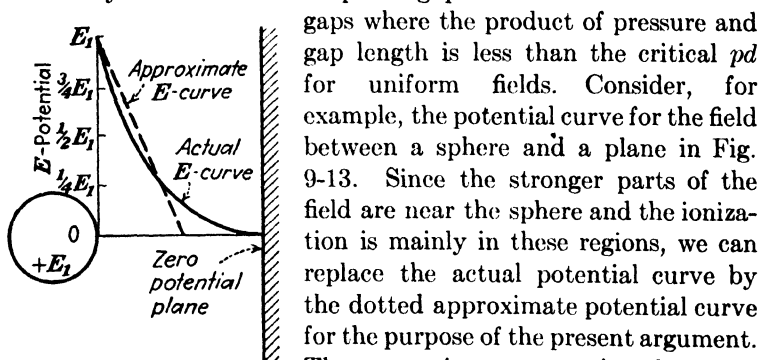


FIG. 9-13.

gaps where the product of pressure and gap length is less than the critical pd for uniform fields. Consider, for example, the potential curve for the field between a sphere and a plane in Fig. 9-13. Since the stronger parts of the field are near the sphere and the ionization is mainly in these regions, we can replace the actual potential curve by the dotted approximate potential curve for the purpose of the present argument. The approximate curve is taken as a straight line; and so we have an equivalent uniform field, which extends only part way across the gap. By this rough analysis, we have reduced the problem to that of the breakdown of a shorter gap sustaining a uniform field. If we recall Fig. 9-4, we note that, if the pd product is below the critical, then a higher voltage is necessary for a self-maintaining discharge with the shorter gap. Of course, at pd 's

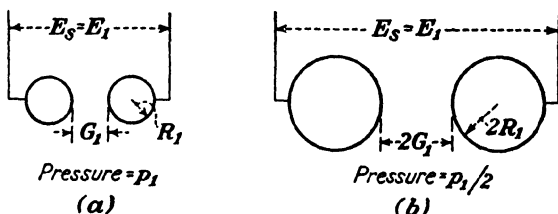


FIG. 9-14.—Illustrating the principle of similitude.

above the critical, the fact that the equivalent gap is shorter means that the sparking potential will be lowered.

There is a general relation between the sparking potential and the pressure in nonuniform fields known as the *principle of similitude*. This principle states that for geometrically similar systems the sparking potential is a function only of the product of gas density and one of the linear dimensions of the system.

This is illustrated in Fig. 9-14, which shows two geometrically similar systems having equal sparking potentials. All the dimensions in Fig. 9-14*b* are double those in Fig. 9-14*a*, and the pressure is halved; therefore, on the basis of the same temperature, the product of gas density by a linear dimension remains constant, and thus the sparking potential remains the same. It should be noted that the principle of similitude applies to the sparking and corona voltages but does not apply to the *breakdown* voltage when corona precedes breakdown. It can easily be seen that this principle reduces to Paschen's law in the special case of uniform fields.

9-6. Sparking and Corona between Concentric Cylinders and between Parallel Wires in Air.—The discharge between con-

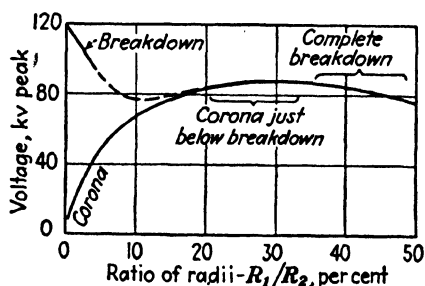


FIG. 9-15.—Corona and breakdown voltages between concentric cylinders in air at atmospheric conditions, tested with a 60 cycle/sec. wave. R_1 = inner radius, R_2 = outer radius = 6.67 cm.

centric cylinders may take the form of corona or a complete breakdown, depending on the size of the inner cylinder or wire relative to the gap length. The results of a typical experiment¹ using alternating potentials are shown in Fig. 9-15. These results are similar to those given for spheres in Fig. 9-10 in that the corona discharge is observed when the radius of curvature of the inner cylinder is small compared with the gap length.

Calculations of the electric intensity at the surface of the wire at the corona or sparking potential show that the intensity for a given size of wire is independent of the size of the outer cylinder, provided that the gap between the wire and cylinder is not too small. The electric intensity at the sparking potential varies considerably, however, with the radius of the inner wire. The results obtained by different observers for air have been collected

¹ General reference 9-2, p. 115

by Schumann,¹ and the resulting curve is shown in curve *AA* in Fig. 9-16. An empirical relation given by Peek to represent the curve *AA* is the following:

$$F = 3,100 + \frac{95.5}{\sqrt{R_1}} \quad (9-5)$$

The effect of varying air pressure and temperature was taken into account by Peek by the introduction into Eq. (9-5) of the

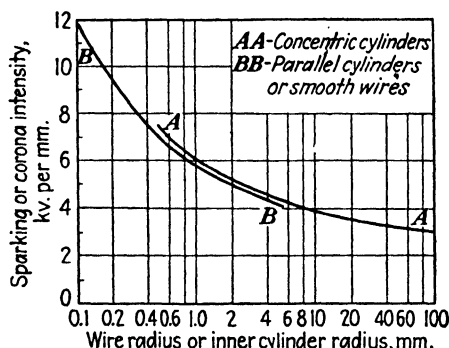


FIG. 9-16.—Sparking intensities at the surface of circular cylinders or wires as a function of the wire radius, based on a.c. tests in air at 760 mm. Hg and 25°C. (General reference 9-3, p. 82.)

relative air density δ' . Then the intensity needed to initiate a discharge is given by

$$F = 3,100\delta' + 95.5 \sqrt{\frac{\delta'}{R_1}} \quad (9-6)$$

and δ' was referred to unity at $p = 760$ mm. Hg and $T_c = 25^\circ\text{C}.$; i.e.,

$$\delta' = \frac{0.392p}{273 + T_c} \quad (\text{mm. Hg, deg. C.}) \quad (9-7)$$

The results obtained by Peek for parallel wires are very similar to those obtained for concentric cylinders. The curve *BB* in Fig. 9-16 shows the intensity for the formation of corona on parallel wires. The voltage between wires required to produce corona can be calculated from the values of intensity in Fig. 9-16 by means of the approximate relation

$$E = 2R_1F \log \frac{G}{R_1} \quad (9-8)$$

¹ General reference 9-3, p. 82.

where G is the distance between wire centers and R_1 is the radius of the wires. This relation holds to an accuracy of 2 per cent for smooth round wires when the ratio G/R_1 is over 100.

The typical corona discharge occurs at all pressures down to one-tenth atmospheric, but at lower pressures the discharge is more diffuse and begins to have noticeable features of the glow discharge described in Chap. X. At atmospheric pressure in air, there is a striking difference between the corona on a positively charged wire and that on a negative wire. The corona on the positive wire is uniform over the whole surface of the wire, whereas that on a negative wire appears in beads or tufts at various points on the wire.

In the case of the wire in a cylinder arrangement, there is a remarkably small difference in the voltage required to start the corona when the wire is positive compared with that required when the wire is negative. This fact, by itself, indicates that the γ -mechanism is not the only one active, for this process of secondary-electron liberation would result in large polarity differences in the corona voltage.

On the basis of evidence obtained in observations on coronas on points Loeb¹ has been led to attach considerable importance to localized disturbances in the gap that lead up to the corona formation. The first predischARGE disturbance is an electron "avalanche" which consists of all the electrons that result from one initial electron by the process of ionization by collision. The presence of these avalanches has been demonstrated by the cloud-chamber method by Raether.²

When the wire is negative, these avalanches can start only in the strong-field region near the wire. The cloud of electrons recedes from the wire, leaving another cloud of slow positive ions behind. The field intensity at the advancing tip of the avalanche is decreasing rapidly as it recedes from the wire; and so the ionization by collision may stop, and the avalanche may die out. This tendency is increased by the space-charge effect of the slow positive ions left behind the avalanche. When the positive ions strike the negative wire, enough secondary electrons may be liberated to maintain the process, and the discharge may spread out over a small area of the wire. On this

¹ General reference 9-7.

² *Z. Physik*, **110**, 611, 1938.

theory, the beaded appearance of the negative corona is ascribed to a variation of γ over the surface, with the discharge taking place where γ is high.

With a positive wire the electron avalanches are directed toward the wire, *i.e.*, into the strong-field region of the gap. However, initial electrons some distance away from the wire cannot start an avalanche because the field is too weak. Therefore, we must look for a mechanism that can *start* with an avalanche near the wire but that can account for a propagation of the

discharge out into the gap. The secondary electrons needed to extend the process outward can arise by photoionization of the gas atoms¹ by short-wavelength radiation from the ionized and excited atoms in the first avalanche. The formation of avalanches farther from the positive wire is facilitated by the positive space charges remaining from the initial avalanche, for these will strengthen the field farther away from the wire.

The steps in the process are sketched schematically in Fig. 9-17. At (a) the first avalanche is nearly complete, and at (b) the

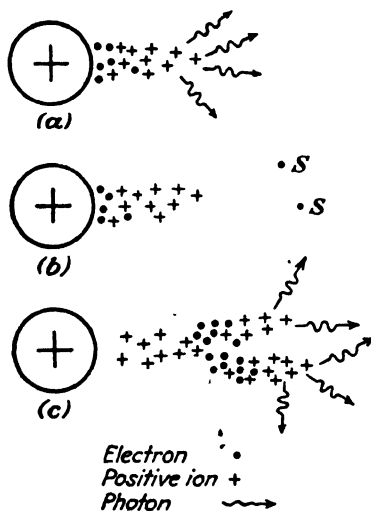


FIG. 9-17.—Formation of electron avalanches.

photons have produced one or two favorably placed secondary electrons *S*. These, as in (c), produce additional avalanches headed toward the wire, and in this way a self-maintaining corona discharge can start. In this instance the secondary processes active are the photoelectric effect in the gas and the distorting effect on the electric intensities caused by the positive space charges.

9-7. Corona Loss on Transmission Lines.—A phenomenon of great practical importance when corona appears on high-voltage transmission lines is the accompanying loss of power. For a.c. lines the corona discharge forms during each half cycle of the

¹ This process would be found particularly in mixed gases where the radiation from one type of molecule has a large enough energy to ionize the second type.

impressed voltage. If the voltage is only slightly above the corona-forming voltage, then the corona discharge will occur only near the peaks of the voltage wave. The accompanying corona current will cause a loss of power from the line, for the instantaneous power $e \times i$ is positive throughout the time of the corona discharge. Peek¹ has conducted extensive studies of the corona loss, and his text should be consulted for detailed information. The general form of the empirical formula developed by Peek for the corona loss from parallel wires is the following:

$$P = k(f + 25)(E - E_d)^2 \quad (9-9)$$

where P , the power loss, is usually expressed in kilowatts per kilometer of conductor, f is the frequency, E is the applied voltage, E_d is the *disruptive critical voltage*, and k is a constant depending on the spacing and size of the wires and on the units used. The disruptive critical voltage E_d is not necessarily the same as the visual corona voltage, but E_d is chosen to give the best agreement of calculated and observed power loss. Peek gives auxiliary formulas for calculating E_d . A comparison between the observed power loss for a two-wire line and the loss calculated by Eq. (9-9) is shown in Fig. 9-18. The observed and calculated values agree quite well except for the voltages near the visual corona voltage of the line. This is unfortunate, for the region of most importance in the design of high-voltage lines is that in the neighborhood of the corona-forming voltage. Therefore, in the design of important extra high-voltage lines, special experimental tests are conducted to test the corona-loss characteristics of the proposed line.²

An alternative formula for the power loss has been developed by Holm.³ This derivation is based on a consideration of the

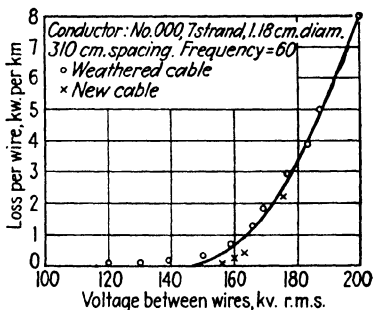


FIG. 9-18.—Corona loss for a two-wire experimental line. Curve shows calculated values. (General reference 9-2, pp. 174, 297.)

¹ General reference 9-2.

² For example, see *Trans. A.I.E.E.*, **52**, 55, 1933 and *Elec. Eng.*, **54**, 498, 1935.

³ *Wiss. Veröff. d. Siemens-Konz.*, **4**, 14, 1925.

movement of the space charges set up by the corona discharge. A simplified form of Holm's equation is as follows:

$$P = kfE(E - E_1) \quad (9-10)$$

in which the symbols have the same meanings as in Eq. (9-9) except that E_1 is the visual corona voltage. Holm's equation has a better theoretical foundation than Peek's but does not predict the power loss with appreciably better accuracy.

Problems

1. (a) If electrodes having the shape of those shown in Fig. 9-1 are moved close together in air at atmospheric pressure and at 20°C., at what spacing would the sparking potential be 400 volts? (b) If the electrodes are moved closer together than in part a, what changes will occur in the sparking potential?
2. What is the sparking intensity in a 12-mm. gap between parallel plates in hydrogen gas at a pressure of 100 mm. Hg and a temperature of 20°C.?
3. At what temperature would the sparking potential for an air gap of 2 mm. at atmospheric pressure be 6,000 volts? Assume that the field is uniform.
4. A concentric-cylinder arrangement of electrodes of given dimensions is to be used in air at a pressure of 100 mm. Hg. How could the sparking potential for these electrodes be predicted by a test at atmospheric pressure?
5. At what peak voltage would corona be expected on a two-wire line consisting of No. 6 A.W.G. solid wires spaced 36 in. apart? The pressure may be taken to be 760 mm. Hg and the temperature 25°C.

General References

- 9-1. THOMSON, J. J., and G. P. THOMSON: "Conduction of Electricity through Gases," Vol. II, pp. 470-573, Cambridge University Press, London, 1933.
- 9-2. PEEK, F. W., JR.: "Dielectric Phenomena in High-Voltage Engineering," 3d ed., Chaps. IV, V, and VI, McGraw-Hill Book Company, Inc., New York, 1929.
- 9-3. SCHUMANN, W. O.: "Elektrische Durchbruchfeldstärke von Gasen," Verlag Julius Springer, Berlin, 1923.
- 9-4. WHITEHEAD, S.: "Dielectric Phenomena (Electrical Discharges in Gases)," Ernest Benn, Ltd., London, 1927.
- 9-5. LOEB, L. B.: "The Problem of the Mechanism of Static Spark Discharge," *Rev. Mod. Phys.*, **8**, 267-293, 1936.
- 9-6. LOEB, L. B., and A. F. KIP: "Electrical Discharges in Air at Atmospheric Pressure," *J. App. Phys.*, **10**, 142-160, 1939.
- 9-7. LOEB, L. B.: "Fundamental Processes of Electrical Discharge in Gases," Chap. X, John Wiley & Sons, Inc., New York, 1939.
- 9-8. VON HIPPEL, A.: "Der elektrische Durchschlag in Gasen und festen Isolatoren," *Ergebnisse Exakten Naturwiss.*, **14**, 79-129, 1935.

CHAPTER X

THE GLOW DISCHARGE¹

10-1. Introduction.—As described at the beginning of the last chapter, the onset of a self-maintained discharge, when the sparking potential is reached, is characterized by a glow discharge, an arc discharge, or a corona discharge, depending on the physical conditions under which the experiment is performed (see Figs. 9-1, 9-2, and 9-3). The term *glow discharge* is used by many writers to describe the whole class of discharges requiring a voltage for maintenance higher than the minimum sparking potential. Others exclude corona discharges from this category. This latter view has been adopted in this book. All discharges characterized by a voltage necessary for maintenance less than the minimum sparking potential will be classed as arc discharges and discussed in the next chapter. Corona discharges have already been discussed, and so this chapter will be devoted to the glow discharge. Since this type of discharge is usually encountered at pressures below atmospheric, the discussion will apply particularly to low or moderate gas pressures, unless otherwise stated.

When an increasing voltage is applied to a pair of plane-parallel electrodes immersed in a gas, the field distribution remains uniform until the sparking potential is reached. With the formation of a self-maintaining discharge the current tends to increase without limit. Actually, since the positive and negative ions have a finite mobility, space charges soon develop and limit the current to a value determined partly by the external circuit resistance and partly by the voltage drop in the discharge itself.

The glow discharge itself is very striking in its display of light, dark, and colored regions, and at low pressures the difference in the appearance of the gas at various points between the electrodes becomes very clearly marked (see Fig. 10-1). It has been customary to classify the various parts of the discharge according

¹ General references for Chap. X will be found on page 356.

to the appearance they present to the observer. This is very convenient; for it happens that the various light and dark regions have particular physical characteristics of their own, and by this method of classification it is easily possible to visualize the region under discussion. A typical glow in air in a cylindrical glass tube might have the appearance sketched in Fig. 10-2. In

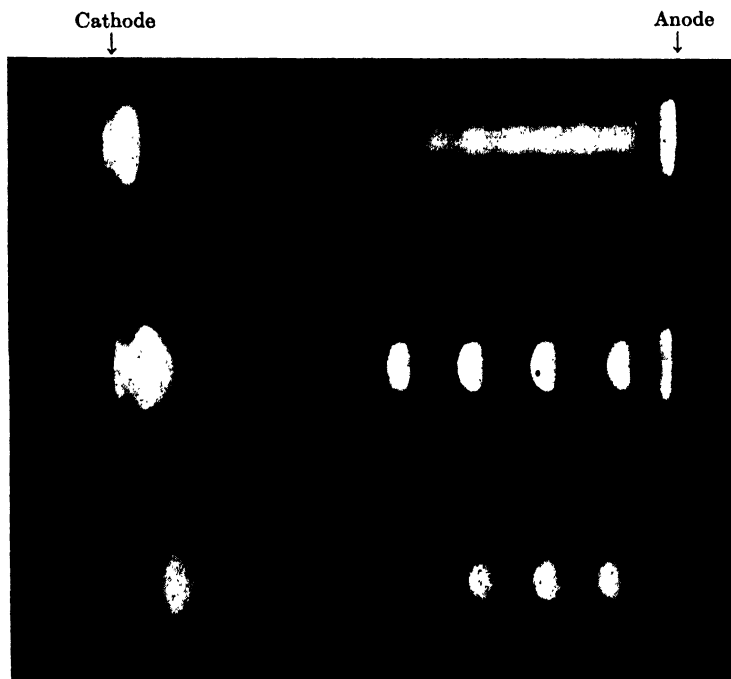


FIG. 10-1.—Glow discharge in air at successively lower pressures (top to bottom).

this figure the parts of the discharge have been labeled, and the color that ordinarily appears in air is indicated. Just below the sketch of the glow is a diagram (Fig. 10-2b) of the potential distribution that might be encountered in this typical case, both before and after the formation of the self-maintaining glow discharge. In Fig. 10-2c the electric-field distribution in the two cases is also represented. From the figure, it is evident that there are four principal parts to the discharge. At the cathode, we find the cathode, or Crookes, dark space. This is the region of most rapidly varying voltage and consequently the region of greatest electric intensity. It is not really a dark space in the

sense that there is a complete lack of luminosity; for there is usually a very thin and velvety glow on the surface of the cathode and a very slight glow throughout the whole region (see Fig. 10-1), but the phenomena taking place in the various parts of the cathode dark space are of such similar character that customarily the region is not subdivided.

Next to the cathode dark space is a luminous region usually called the cathode or negative glow. This is a region of very

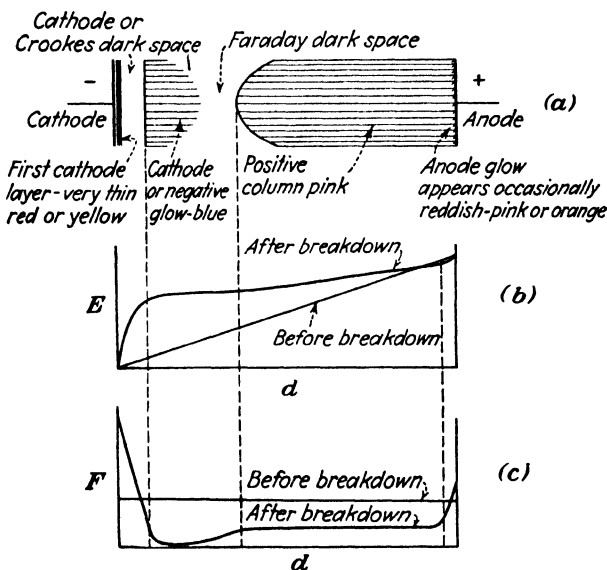


FIG. 10-2.—Potential and electric intensity distribution in the glow discharge.

low field strength and consequently practically constant potential. In some discharges the electric-intensity curve actually passes through zero and becomes negative in the negative-glow region. Under such conditions the potential curve passes through a maximum and drops slightly, and the flow of current is for a short distance counter to the direction of the electric field.

The negative glow fades off into another dark region called the Faraday dark space where the intensity and potential curves again begin to rise. This, in turn, is followed by another luminous region, the positive column, extending all the way to the anode. Under some conditions the positive column may be striated and may consist of a succession of glowing disks separated

by short dark spaces (Fig. 10-1). In the uniform positive column the field intensity is quite low and practically constant. The voltage curve consequently rises with constant slope as the anode is approached. In long tubes the positive column makes up the greater part of the discharge; for the space occupied by the cathode dark space, negative glow, and Faraday dark space is only a few centimeters and is entirely independent of the length of the tube. Regardless of this, the positive column seems to be the least necessary part of the glow, as far as maintenance of the discharge is concerned. As long as the current is kept constant, increasing or decreasing the electrode separation in the tube

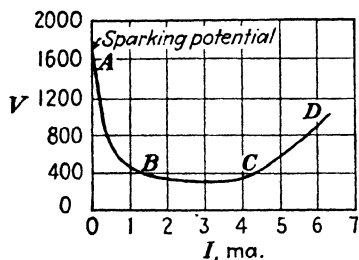


FIG. 10-3.—Volt-ampere characteristic of a glow discharge.

results only in increasing or decreasing the length of the positive column, unless the anode-cathode separation is shorter than the space ordinarily occupied by the other parts of the discharge. If the cathode is moved about in a large vacuum chamber, the anode remaining fixed, it is a characteristic of the glow discharge that the cathode dark space, negative glow, and Faraday dark space move about with the cathode while the positive column fills the rest of the tube, lengthening and shortening as the anode-cathode separation varies.

We have said in the preceding paragraph that the positive column fills the balance of the tube not occupied by the other three parts of the discharge. The statement is certainly true at the low pressures at which stable glow discharges are usually encountered. However, as the gas pressure is increased, a condition is eventually reached when the positive column begins to shrink away from the tube walls and occupy a path of small radius near the axis of the tube. When this occurs, the pressure is always found to be high enough and the mean free path of the electrons and positive ions short enough for an appreciable number of ions to be lost by recombination in the gas before they reach the tube walls.

Figure 10-3 is a sketch of a typical voltage-current curve for a glow discharge when the pressure and electrode separation are

such that the corresponding value for pd is on the right of the minimum of the sparking potential curve (see Fig. 9-4). As the voltage between the electrodes is raised, the voltage distribution is determined only by the electrode configuration until the sparking potential is reached.¹ At the sparking potential the non-self-maintaining Townsend discharge changes to a self-maintaining discharge, and the current begins to increase rapidly. Space charges develop because of the low mobility of the positive ions formed in the discharge, and a large part of the tube voltage appears across the positive space charge forming at the cathode surface. With the distortion of the electric field, it becomes possible to have a self-maintained discharge at a lower voltage than initially; consequently, the voltage drop across the tube decreases. As the current continues to increase, the electric field becomes more distorted and the voltage necessary to maintain the discharge drops still more. This process continues until the potential drop across the positive space charge at the cathode is approximately equal to the minimum sparking potential of the gas. The potential can decrease no further; for, as we shall see in the next section, the conditions for a self-maintained discharge must be satisfied in the cathode dark space, and a certain minimum potential is necessary for this to occur.

Ordinarily, the characteristic shown in Fig. 10-3 cannot be traced out completely. In most cases the external circuit resistance is low enough so that the portion of the curve between *A* and *B* is traversed very rapidly.² When the impressed voltage reaches the sparking potential, there is a sudden breakdown into a glow discharge. When this occurs, the voltage drops and the current increases almost instantaneously to the value corresponding to point *B*. Because of the voltage drop in the other parts of the discharge, the voltage at *B* is always somewhat higher than the minimum voltage on the sparking potential curve.

If the current is sufficiently small or if the cathode is of sufficient surface area, at *B* the negative glow will have collapsed and will cover only a portion of the cathode surface. Further

¹ Only in the case of large parallel-plate electrodes would the voltage distribution be uniform and the field intensity constant.

² For the effect of circuit resistance on the glow characteristics, see Sec. 10-3, page 245ff.

increases in current will then be accompanied by an increase in the area of the cathode covered by the negative glow and a slight decrease in tube voltage. This is due to the fact that although the voltage drop across the cathode dark space remains constant until the cathode is completely covered by the negative glow, the positive column always has a falling volt-ampere characteristic. When the negative glow completely covers the cathode surface, the tube voltage begins to rise again (indicated by *C*, Fig. 10-3), for then the voltage drop across the cathode dark space increases rapidly with an increase in current and completely overshadows the slight voltage decrease that occurs in the positive column.

How is the appearance of the discharge affected by the change in current? We have already mentioned that until the point *C* is reached the area of the cathode covered by the negative glow increased. However, the appearance of the rest of the discharge remains practically the same. When the cathode is completely covered, a further increase in current is accompanied by a decrease in the thickness of the cathode dark space and an increase in the thickness of the negative glow. At the same time the positive column gets shorter and brighter, and the Faraday dark space thus moves continuously nearer the anode.

Let us now examine each of these four divisions of the glow discharge in more detail.

10-2. The Cathode, or Crookes, Dark Space.—As was mentioned in the preceding section, this region is not strictly a dark space as the name implies. It is pervaded throughout by a faint luminosity which sometimes becomes brighter as the cathode is approached and may develop into a definitely bright region at the cathode surface, called the first cathode layer. The difference in brightness, however, between the cathode dark space and the negative glow is so very marked in most gases that there is no difficulty in distinguishing its boundary. The boundary in oxygen, for example, is so well marked that it is possible to fix its position within a fraction of a millimeter. In other gases, the dark space is not so well defined visually, and electrical methods must be used to determine the position of the dark-space boundary. Since the cathode dark space is the last to disappear as the anode-cathode separation in a glow discharge is decreased, a method (aside from the visual one) of determining the dark-

space thickness would be to move the anode of the discharge closer and closer to the cathode. If this is done, it is found that the potential difference required to maintain a given current decreases uniformly as the positive column disappears. This corresponds to the region *AB* of Fig. 10-4. Then, if an anode glow is present, there is a sudden drop in voltage (*B* to *C*) as that disappears. Since the Faraday dark space and the negative glow are regions of practically constant potential, the electrode voltage remains essentially constant as the anode is moved through these regions. But when the anode reaches the edge of the cathode dark space at *D*, the voltage necessary to maintain the discharge rises very rapidly; and unless the circuit voltage is high the glow is soon extinguished. This method furnishes a quite critical determination of the dark-space thickness.

The cathode dark space is a region of high positive space charge. Consequently, we should expect the electric intensity in the cathode dark space to be nonuniform. Aston¹ investigated the field strength near the cathode by sending a beam of electrons through the region and observing the deflection of the spot produced on a fluorescent screen. He found that the distribution was very close to linear, as shown by the solid curve in Fig. 10-5. The field strength is very high at the cathode and falls to zero at the boundary of the dark space. This, then, is another way in which the thickness of the dark space has been determined. More recent experiments, making use of spectroscopic means to investigate the field distribution near the cathode, seem to indicate that, although the electric-intensity distribution is approximately linear over most of the dark space, it reaches a maximum a short distance from the cathode surface and then decreases slightly very much as indicated by the dotted curve in Fig. 10-5. It is easy to see from Fig. 10-5 that the maximum field strength in the cathode dark space is not the ratio of the fall in potential across the dark space to the length of the

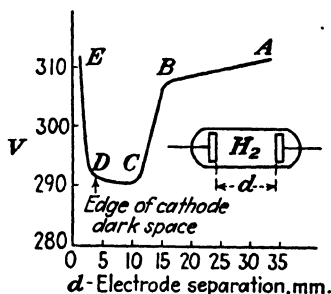


FIG. 10-4.—Effect of electrode separation on the voltage drop in the glow. (A. Guntherschulze, *Z. Physik*, 30, 175, 1924.)

¹ *Proc. Roy. Soc. (London)*, A84, 526-535, 1911.

dark space as is sometimes stated in discussions of glow discharges. Actually, the maximum field strength is approximately twice that value and occurs right at, or close to, the cathode surface. One must not make the mistake, however, of supposing that it is the high field pulling electrons from the cathode that initiates or maintains the discharge. The maximum field strengths encountered (10^5 to 4×10^5 volts/m. depending upon the pressure) are far short of the values necessary for such an explanation. One must look farther than that for a possible explanation of the maintenance of the glow.

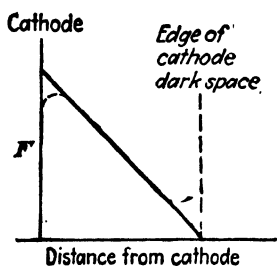


FIG. 10-5.—Electric-intensity distribution in the cathode dark space.

Three experimental facts stand out in any investigation of conditions in the cathode dark space:

1. The cathode dark space is the most important part of the glow discharge. The discharge can exist without a positive column, Faraday dark space, or negative glow; but it cannot exist without a cathode dark space. When the electrodes are moved closer together than the normal thickness of the cathode dark space, the voltage necessary to maintain the discharge rises very rapidly.

2. The lowest values of cathode fall occur, in any gas, with cathode materials that most readily emit electrons. Thus, it is possible to have ratios of 2 or 3 to 1 between the values for such metals as platinum and potassium. In general, the alkali and alkaline earth metals are associated with the lowest cathode falls.

3. With any metal the lowest values of cathode fall occur with the lighter of the noble gases. The most favorable combination of all for bringing down the cathode fall would be the use of helium, neon, or argon with an alkali metal. In such a case, it is possible to have a value of cathode fall of as little as 64 volts, whereas with such a combination as air and platinum or iron the cathode fall may exceed 400 volts.

These facts suggest a possible mechanism for the maintenance of the glow. Since the glow is a self-maintaining discharge and needs only the presence of a cathode dark space in order to be maintained, the conditions necessary for a self-maintaining discharge must be satisfied in the cathode dark space. This

means that, if we adopt the mechanism for a self-maintaining discharge that was presented in Sec. 8-5, each electron leaving the cathode must initiate enough ionizing processes in crossing the dark space so that the positive ions so formed, in returning to the cathode, are sufficient or possess enough energy to liberate another electron. On this basis, then, since the work function of a metal is the work required to extract an electron from the metal, we should expect that, on the average, with metals of high work function more positive-ion impacts would be necessary to eject an electron than with low-work-function materials. If a large number of positive ions are necessary to produce a secondary electron, the electron must in turn produce a large number of positive ions by collision in the dark space in order to have the discharge maintained. All this involves a large expenditure of energy and consequently a large cathode fall of potential. On the other hand, a material of low work function would yield a secondary electron after fewer collisions of positive ions, and this electron would have to produce fewer ionizing collisions in the dark space in order to keep the discharge going. In this case, we should expect a low cathode fall of potential. Thus we see that, in any given gas, the cathode fall should be roughly proportional to the work function of the cathode material. That this is so, at least in a qualitative way, can be seen from Table 10-1 below in which the normal cathode fall of poten-

TABLE 10-1

Cathode material	Work function ϕ of cathode material, volts	Calculated cathode fall $V_n = 57.35\phi$	Experimental value of cathode fall
Magnesium.....	2.7	155	153
Aluminum.....	3.0	172	170
Zinc.....	3.4	195	184
Tin.....	3.8	218	226
Silver.....	4.1	235	216
Carbon.....	4.34	249	230

tial with several different cathode materials in hydrogen is compared with the value calculated on the basis of a direct proportionality between the cathode fall and the work function of

the cathode material. Obviously, this proportionality does not hold true in the case of all cathode materials, but there is a general correspondence between the work function and the normal cathode fall of potential. The discrepancies that occur are partly due to uncertainties in the experimental values of the work function and the normal cathode fall of potential and partly

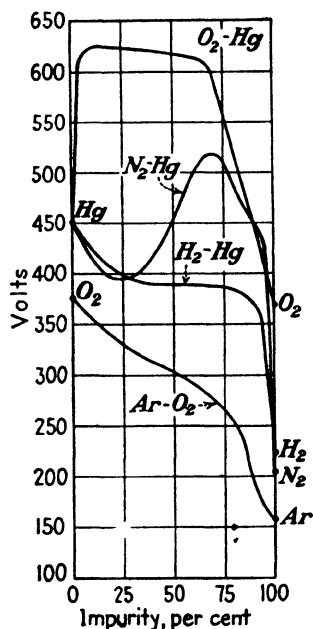


FIG. 10-6.—Effect of impurities on the normal cathode fall of potential. (A. Guntherschulze, *Z. Physik*, **21**, 50, 1924.)

gases all have important metastable states, it is possible that these long-lived excited atoms produce an appreciable fraction of the secondary electrons by collisions with the cathode. These factors would all tend to reduce the cathode fall in the noble gases; for with an easily ionized gas or with a high probability of the presence of metastable states, less energy need be expended in producing enough ionization so that the discharge can be maintained.

The dependence of the normal cathode fall of potential on electrode material and type of gas used is illustrated by the

due to the fact that the actual ionizing and deionizing processes active are probably much more complex than the approximate theory outlined above would seem to indicate. As we shall see a little later, the cathode fall can vary over quite wide limits when even very slight amounts of impurities are present in the discharge.

The effect of various gases on the cathode fall is also evident when we compare the ionizing ability of the electrons in the various gases. Figure 8-9 gives us a hint of this. Since over a large portion of the dark space, and especially near the boundary of the negative glow where most of the ionization occurs, F/p has a value less than 10^4 V/m, mm. Hg, α/p is considerably higher in argon and helium than in air or hydrogen. This situation is true for all the noble gases. Furthermore, since these

following table. As was mentioned before, the values in this table cannot be taken as accurate; for, in any case, there are many factors that may enter to cause the results to be slightly in error. The effect of impurities in the gas might be taken as an illustration. Figure 10-6 shows the wide limits through which the cathode fall may vary when even slight amounts of an

TABLE 10-2.—NORMAL CATHODE FALL, VOLTS, IN VARIOUS GASES*

Metal	Air	O ₂	N ₂	H ₂	He	Ne	Ar	Hg
Na	~ 200	...	178	185	80	75	...	...
K	~ 180	...	170	94	59	68	64	...
Cu	370	...	208	214	177	220	130	447
Ag	279	...	233	216	162	150	131	318
Mg	224	310	188	153	125	94	119	...
Ca	...	...	157	...	86	86	93	...
Ba	...	...	157	...	86	...	93	...
Zn	277	354	216	184	143	...	119	...
Hg	...	...	226	337	142	...	...	340
Al	229	310	179	170	141	120	100	245
C	...	...	...	280	...	...	...	475
Sn	266	...	216	226	...	...	123	...
W	...	...	...	...	...	125	...	305
Fe	269	290	215	250	150	150	165	298
Ni	226	...	197	211	158	140	131	275
Pt	277	364	216	276	165	152	131	340

* General reference 10-3, p. 103; M. STEENBECK, MÜLLER-POUILLET, "Lehrbuch der Physik," Vol. IV. Part 3, p. 406, Friedrich Vieweg & Sohn, Brunswick, Germany, 1933.

impurity are present. In the case of a mercury discharge, the presence of even 1 or 2 per cent of oxygen will raise the cathode fall from about 450 volts to over 600 volts even though the cathode fall in oxygen is less than in pure mercury vapor. In other cases, such as with hydrogen in a mercury-vapor tube, the cathode fall is lowered by the addition of the impurity. Although there is some evidence that the gas reacts with the metal of the cathode to form complex compounds which alter the cathode fall, there seems to be no set rule governing the effect of mixtures of gases on the cathode fall of potential. This is unfortunate, for the problem of determining the factors influencing the cathode fall of potential is of considerable practical importance now that glow discharges are being extensively used for many engineering applications.

Up to now, we have been tacitly assuming in this discussion that we were dealing with a cathode dark space in the *normal* condition. We have called the cathode fall the *normal* cathode fall and the length of the dark space the *normal* thickness of the dark space. Now, what is meant by a *normal* cathode fall as differentiated from an *abnormal* cathode fall? These names are rather ambiguous, for there is nothing particularly normal or abnormal about the glow discharge. The terms were merely applied to two different conditions in the cathode dark space. In discussing Fig. 10-3, we pointed out that if the current in the glow discharge is sufficiently small (corresponding to point *B*, for example) the cathode will be incompletely covered by the negative glow. As the current is increased, however, the area of the cathode covered by the negative glow increases and the cathode-dark-space thickness remains constant until the cathode is completely covered by the negative glow (at *C*, for example). Then any further increase in current causes the thickness of the dark space to decrease and the voltage across the tube to rise rapidly. We say that a *normal* discharge exists between *B* and *C* when the cathode is incompletely covered by the negative glow, but between *C* and *D* we have an *abnormal* discharge.

In the condition for a *normal* discharge the area of the cathode covered by the glow increases directly with the current so that the current density remains constant. Its value is called the *normal current density*. The thickness of the cathode dark space and its potential drop are also constant with increasing current; they are called, respectively, the *normal cathode-dark-space thickness* and the *normal cathode fall*. The latter has a value approximately equal to the minimum sparking potential and is independent of the pressure. The normal cathode-dark-space thickness is related to the pressure in a way that reminds one very much of Paschen's law, namely,

$$pd_n = \text{constant} \quad (10-1)$$

The constant in Eq. (10-1) is also very nearly equal to the value of pd corresponding to the minimum on the sparking-potential curve. This is a further verification of the statement that the conditions necessary for a self-maintained discharge must be satisfied in the cathode dark space. In general, d_n represents a distance of between 10 and 20 electron mean free paths; depending upon the kind of gas and cathode material.

As shown in Eq. (10-2) the normal current density is proportional to the square of the pressure.

$$J_n = Kp^2 \quad (10-2)$$

Here K is the constant of proportionality. If we consider the gas density to be directly proportional to the pressure, Eq. (10-2) is in agreement with an extension of the law of similitude, discussed in Sec. 9-5. Thus, we can say that in two geometrically similar systems, when the pressure is increased in proportion to the decrease in linear dimensions, not only is the sparking potential the same, but with a given potential difference between electrodes the current that flows is the same. That Eq. (10-2) is in agreement with this principle is evident when we consider that a decrease in all the linear dimensions of a discharge tube in proportion to the increase in pressure involves an increase in the current per unit area in proportion to the square of the pressure if the total current is to remain constant.

After the current has been raised to a point where the cathode is completely covered and we pass into the state of an abnormal discharge, the cathode fall increases very rapidly with increase of current or decrease of pressure. The thickness of the dark space, on the other hand, decreases with the increase in current or pressure. Experimentally, the abnormal cathode voltage drop and dark-space thickness have been found to follow quite closely the relations

$$d_a = \frac{A}{p} + \frac{B}{\sqrt{J_a}} \quad (10-3)$$

$$V_a = C + \frac{D \sqrt{J_a}}{p} \quad (10-4)$$

In these expressions, V_a is the abnormal cathode fall of potential in volts, d_a is the abnormal dark-space thickness in meters, J_a is the abnormal current density in amperes per square meter, p is the pressure in millimeters of mercury, and A , B , C , and D are constants. The order of magnitude of these constants can be ascertained from the fact that in the case of a hydrogen discharge with aluminum electrodes the values of A , B , C , and D are approximately 0.0023, 0.004, 144, and 2.55, respectively.¹ It is worth noting again at this time that, since the experimental conditions

¹ SLEPIAN, J., and R. C. MASON, *Elec. Eng.*, **53**, 515, 1934.

under which any of these constants are derived are extremely difficult to reproduce, the constants and the equations to which they

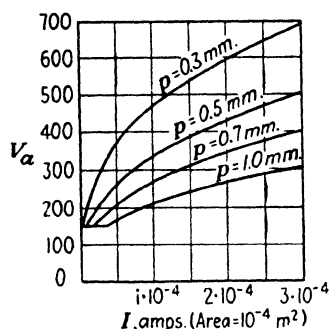


FIG. 10-7.—Dependence of cathode fall on current (neon discharge—iron cathode). (Derived from data of A. Guntherschulze, *Z. Physik*, **59**, 433, 1930.)

refer can be taken only as indicative of general trends. Consequently, they will give only approximate values for any of the variables.

Figures 10-7, 10-8, and 10-9 show the dependence of the abnormal cathode fall of potential on the gas pressure and the current as determined experimentally for a discharge between iron electrodes in neon. The horizontal portions of the curves in Figs. 10-7 and 10-8 and the small circle in Fig. 10-9 correspond to conditions in the normal cathode dark space.

Before leaving the discussion of this part of the discharge, it may be well to review briefly the principal facts gained from our survey. The cathode, or Crookes, dark space is the most important part of the glow discharge, for all other parts of the discharge depend upon it for their maintenance. The discharge as a whole is self-maintaining, with a voltage greater than the minimum sparking potential. A single electron leaving the cathode makes enough ionizing collisions in passing through the dark space so that all the positive ions for which it is responsible, both directly and indirectly, will, in falling into the cathode, produce another electron and thus keep the discharge going. The voltage necessary to maintain the process of electron generation by secondary emission from the cathode is approximately equal

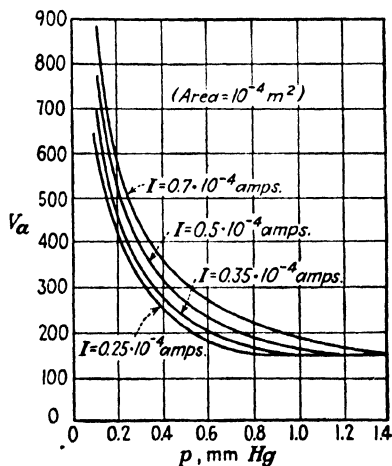


FIG. 10-8.—Dependence of cathode fall on pressure (neon discharge—iron cathode). (*Loc. cit.*)

to the minimum sparking potential and over a wide range is independent of electrode separation, current density, and pressure, depending only on electrode material and the nature of the gas. However, when the total current becomes so large that the cathode is completely covered with the negative glow, the voltage drop across the cathode dark space increases rapidly. Just as V_n is constant, so is d_n , the thickness of the cathode dark space, until the cathode is completely covered with the negative glow, and then it decreases with a further increase of current.

10-3. The Negative Glow and the Faraday Dark Space.—As shown in Fig. 10-2 the negative glow is that luminous portion of the glow discharge adjacent to the cathode dark space. It usually has a distinct boundary only on the side toward the cathode. On the anode side, it fades off gradually into the void of the Faraday dark space. Since there is no sharp line of demarcation and since the phenomena encountered in these two regions are so closely related, it is convenient to discuss them together.

In the last section, it was shown that the conditions necessary for a self-maintaining discharge are satisfied in the cathode dark space. The electrons from the cathode, however, do not lose all their energy in ionizing collisions in the dark space. When they leave, they still possess sufficient energy for an appreciable amount of ionization and excitation. This, then, suggests an explanation for the dependence of the length of the negative glow on the gas pressure and the cathode fall of potential. Since the length of the negative glow increases as the pressure decreases and as the cathode fall increases in an abnormal discharge, it must be dependent upon the distance the high-speed electrons coming from the cathode can travel before losing so much energy that they have not enough left to produce either ionization or excitation. Beyond this point, in the Faraday dark space, they have very little energy; and the current must be

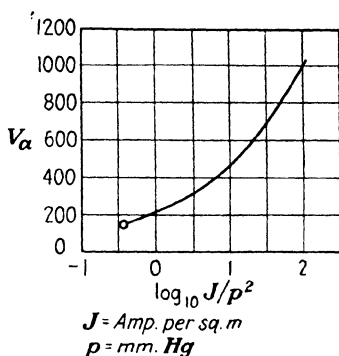


FIG. 10-9.—Dependence of cathode fall on current density and pressure (neon discharge—iron cathode). (*Loc. cit.*)

carried by the diffusion of electrons from regions nearer the cathode.

In the region of the negative glow and the Faraday dark space, the space potential is practically constant and in many discharges actually decreases slightly. Since the electric intensity is given by the gradient of the space potential, this would imply that over a portion of the region in some discharges the electric intensity is actually in such a direction as to retard the flow of electrons toward the anode. This situation is illustrated in a somewhat exaggerated form in Fig. 10-10 where the abscissas

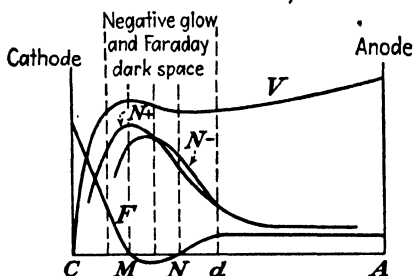


FIG. 10-10.—Variation of ion density in the negative glow and Faraday dark space.

represent distance as measured from the cathode. Under the circumstances illustrated in the figure a current is possible through the tube only if the electrons can diffuse against the electric field in the region MN . This in itself suggests that the density of electrons in the negative glow is very large. Measurements of the electron density in the negative glow by means of probes have verified this. Equation (6-30), however, tells us that in any space-charge region

$$\frac{d^2E}{dx^2} = -\frac{\rho}{\epsilon} \quad (6-30)$$

or that, since $F = -dE/dx$,

$$\frac{dF}{dx} = \frac{\rho}{\epsilon} \quad (6-30a)$$

Therefore, since ρ is the net space-charge density and dF/dx is small in the negative glow, the density of positive ions must be nearly the same as the density of electrons, as indicated in Fig. 10-10. According to Eq. (6-30a) the density of positive ions should exceed the density of electrons throughout the cathode

dark space; but in the negative glow they should be approximately equal, and in the Faraday dark space there should be a net negative space charge. There is no abrupt change in the density of either positive ions, N_+ , or negative ions, N_- , on passing from the Faraday dark space into the positive column. The densities simply decrease and approach equality.

A qualitative explanation of the high densities of positive and negative ions in this region is evident when we consider that the ionizing ability of electrons is relatively large even when they have energies of the order of 20 or 30 volts (see Fig. 8-4) and their maximum ionizing ability occurs in the range between 50 and 150 volts. Thus, even though the energy of the electrons leaving the cathode dark space is only a fraction of the energy that an electron would have after traversing the whole cathode dark space without collisions, it is still large enough for the electrons to produce considerable ionization by collision in the negative glow. Furthermore, the field strength in the negative glow and the Faraday dark space is very low; consequently, there is no reason for the ionization to disappear except by diffusion or recombination. Since most of the negative ions present are free electrons, recombination is relatively unimportant as a deionizing agent (see Sec. 7-3). Thus, the ion density builds up until the loss in ionization by diffusion is essentially equal to the rate of production of ions.

10-4. The Uniform Positive Column.—The positive column in any discharge is independent of the phenomena occurring at the cathode as long as it is continually supplied with electrons. It is the least important part of the discharge, and yet the most spectacular, occupying all the discharge space except a narrow region near the cathode and giving the discharge its characteristic display of color. Since the width of the cathode dark space, negative glow, and Faraday dark space is determined, in general, by the mean free path of the electrons (*i.e.*, by the conditions necessary for a self-maintaining discharge and by the distance that the electrons from the cathode travel before losing all their energy), the positive column fills most of a long discharge tube; and if the anode-cathode spacing is lengthened or shortened, it is the positive column that suffers the change. The positive column behaves more like a pure resistance, at least as far as variations in length are concerned, than any other part of the

discharge or than any other discharge, for that matter. As is evident from Figs. 10-2 and 10-10, the voltage gradient in the positive column is practically constant. Thus, since the space-charge density is given by

$$\frac{dF}{dx} = \frac{\rho}{\epsilon} \quad (6-30a)$$

we can see that the net space charge is essentially zero throughout the uniform positive column. This means that, although the density of electrons in the positive column usually exceeds 10^{15} per cu. m., there must be an approximately equal number of positive ions per unit volume. The characteristics of the positive column of any discharge (glow or arc) are essentially the same. However, we shall divide the various discharges into two general classes, namely, low-pressure small-current discharges and high-pressure high-current discharges. Only the former will be discussed at this time. A discussion of the latter class will be postponed until the next chapter.

Most of the current in the positive column is carried by electrons, for their average velocity and mobility are much greater than those of the positive ions even when no electric field is present. A positive column, however, cannot exist without a slight potential gradient; for, in the positive column, there is a continual loss of electrons, and only a sufficient number are regenerated to supply the current necessary for the discharge. The loss of electrons occurs principally by diffusion from the center of the discharge, where the greatest density of electrons exists, to the cooler gas around the positive column or to the walls of the vessel. At low pressures the electron mean free path is long, and the discharge tends to fill the whole tube; but at high pressures the discharge may be confined to a small channel down the center of the tube. At the walls or in the cool gas, recombination occurs between the electrons and the positive ions. Diffusion is not the only mechanism, however, by which electrons are lost from the discharge path. The electrons may combine with neutral molecules to form negative ions which, in turn, can easily recombine with positive ions; or the electrons may combine directly with positive ions to form neutral atoms. Diffusion, however, is by far the most important process. Energy must be continually supplied to the gas not only to regenerate these lost

electrons, but also to accelerate those primary electrons which lose energy in ionizing, in exciting luminosity, or in any kind of inelastic collision. Thus, a certain potential gradient is always necessary in the positive column.

Because of the high speed and the long mean free path of the electrons compared with the positive ions, the insulating walls of the confining vessel soon become charged negatively with respect to the gas in the tube. The potential of the wall, however, ultimately adjusts itself—attracting positive ions and repelling electrons—until equal numbers of the two kinds of ion come up to it. Under that condition, the resultant current to the wall is zero, as it must be because the wall is insulating. The gas space

itself cannot acquire either a positive or a negative space charge, for if it did it would immediately attract ions of the other sign until the space charge was neutralized. Because of the fact that the walls of the tube become charged negatively, the equipotential surfaces are not planes

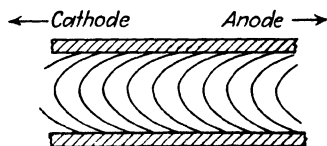


FIG. 10-11.—Equipotential surfaces in the uniform positive column.

perpendicular to the axis of the tube but rather parabolalike surfaces with their convex side toward the cathode, as indicated in Fig. 10-11. It is interesting to note that the end of the positive column next to the Faraday dark space in most glow discharges conforms to the shape of one of these equipotential surfaces.

We are now in a position to determine the relationship between the ordinary coefficient of diffusion of electrons and positive ions and the so-called *ambipolar coefficient of diffusion* which must be applied to calculations of the radial motion of ions when an electric field accelerates the positive ions toward the wall and retards the electrons. This is the coefficient of diffusion that was used in the problem at the end of Sec. 7-2, and it is the coefficient of diffusion that must be used for computing the radial loss of ionization in the positive column of the glow.¹

In the positive column there will be a general motion of ions in a radial direction due to two separate causes. In the first place, the concentration gradient itself will cause a diffusion of ions toward the walls; in the second place, since the potential distribution is as indicated in Fig. 10-11, there will be a motion of ions toward (or away from) the wall due to the radial component of the electric field. Combining these two effects and recalling

¹ This derivation follows that given in general reference 10-3, Vol. I, p. 199.

Eqs. (7-1) and (7-18), we see that the net number of positive ions per second moving across a given small (unit) area parallel to the walls is given by

$$N_+ \bar{V}_+ = -D_+ \frac{dN_+}{dr} + N_+ K_+ F_r \quad (10-5)$$

where N_+ is the density of positive ions, $\bar{V}_+$ is the average drift velocity of the positive ions in a radial direction, D_+ is the coefficient of diffusion for positive ions, K_+ is the mobility of the positive ions, and F_r is the radial component of the electric intensity. An equation exactly like Eq. (10-5) can be written for the electrons, as follows:

$$N_- \bar{V}_- = -D_- \frac{dN_-}{dr} - N_- K_- F_r \quad (10-6)$$

But we know from experiment that the density of positive ions in the positive column is the same as the density of electrons and that in the steady-state condition the average velocity of the two kinds of ion in the radial direction is also the same. Therefore, we can write

$$N_+ = N_- = N$$

and

$$\bar{V}_+ = \bar{V}_- = \bar{V}$$

Now, by combining Eqs. (10-5) and (10-6) so as to eliminate F_r and making the substitutions indicated above, we see that for either kind of ionization

$$N \bar{V} = - \frac{D_+ K_- + D_- K_+}{K_+ + K_-} \frac{dN}{dr} \quad (10-7)$$

Thus, we can say by analogy with Eq. (7-18) that the ambipolar coefficient of diffusion is given by

$$D_a = \frac{D_+ K_- + D_- K_+}{K_+ + K_-} \quad (10-8)$$

For most problems, this can be put in a somewhat simpler form; for, approximately,

$$\frac{K}{D} = \frac{q_e \lambda / mC}{\frac{1}{2} \lambda C} = \frac{3q_e}{mC^2} \quad (10-9)$$

This follows from Eqs. (7-13) and (7-23) when we consider that the radial field strength is very low and when we neglect the small correction factors. But since $\frac{1}{2} mC^2 = (\frac{3}{2}) kT$, we see that

$$\frac{K}{D} = \frac{q_e}{kT} \quad (10-10)$$

Thus, we can rewrite Eq. (10-8) as follows:

$$D_a = \frac{D_+ D_-}{K_+ + K_-} \left(\frac{q_e}{kT_-} + \frac{q_e}{kT_+} \right) \quad (10-8a)$$

When we are considering the motions of electrons and positive ions, however, we always find that $T_- \gg T_+$ and $K_- \gg K_+$. Therefore, Eq. (10-8a) can be simplified to

$$D_a = \frac{D_+ D_-}{K_-} \left(\frac{q_e}{kT_+} \right)$$

or, in view of Eq. (10-10),

$$D_a = D_- \frac{K_+}{K_-} \quad (10-11)$$

Thus, the ambipolar coefficient of diffusion is essentially equal to the coefficient of diffusion of the electrons multiplied by the ratio of the mobility of the positive ions to that of the electrons.

The size of the discharge tube will naturally affect the potential gradient in the positive column greatly. As the problem at the end of Sec. 7-2 indicated, the number of ions produced per unit volume must be increased as the walls are brought closer together if the total ionization in the positive column is to remain constant and if the current through the tube is to remain constant. This means that since ionization by collision is the only source of ions the potential gradient in the positive column must be increased as the tube radius is decreased. Figure 10-12 shows the experimental results obtained in helium with a tube current density of 140 amp./sq. m. The strong dependence of the positive-column gradient on tube radius is evident from the figure.

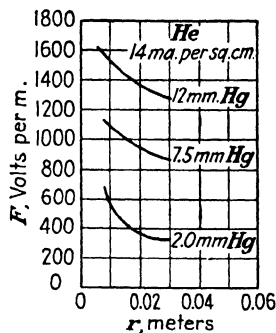


FIG. 10-12.—Variation of positive-column gradient with tube diameter. (A. Lompe and R. Seeliger, *Ann. Physik*, **15**, 300, 1932.)

Figures 10-13 and 10-14 show the variation in the positive-column gradient with the current through the tube and the pressure in the tube, respectively. The falling voltage-current characteristic illustrated in Fig. 10-13 for an argon discharge is typical of all positive columns. It is this property of having a negative dynamic resistance that makes the glow discharge useful in many circuit applications, particularly those involving the production of sustained oscillations (see Sec. 10-9). It also makes it possible to devise tubes, such as voltage-regulator tubes, that have practically a constant voltage drop as long as the current stays within certain limits; for by proper design the

increase in voltage drop at the cathode, as the current increases in the range of an abnormal discharge, can be just compensated by the decrease in voltage drop in the positive column. Figure 10-14 which was obtained from data for an argon discharge with

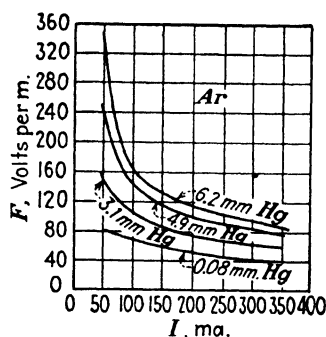


FIG. 10-13.—Dependence of positive-column gradient on current. (*Loc. cit.*)

several different tube currents shows that, except for small values of tube current, the voltage gradient increases only slightly as the pressure increases.

Although enough is now known about the ionizing and deionizing processes active in the positive column to predict its characteristics almost exactly,¹ it is beyond the scope of this volume to go into these calculations. We can, however, justify the curves in Figs. 10-13 and 10-14 in a very approximate and qualitative manner. If the efficiency of ionization remained constant, we might expect the positive-column gradient to be substantially constant with varying current; for the increase in ion density will be approximately proportional to the increase in current, and the principal loss in ionization, diffusion to the tube walls, is proportional to the ion density. Thus, the energy input need increase only in proportion to the current, and the voltage gradient can remain constant. Actually as Fig. 10-13 shows, the voltage gradient decreases with an increase in current; therefore, it is evident that the ion density does not have to increase in proportion to the increase in current in order to maintain an equilibrium condition.

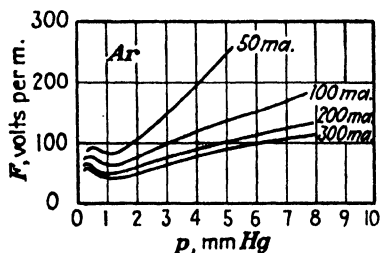


FIG. 10-14.—Dependence of positive-column gradient on pressure. (*Loc. cit.*)

The dependence of the voltage gradient on the pressure can be justified in a similar manner. Since the energy input to the positive column per unit volume per second is FJ and since

¹ See general reference 10-3, pp. 82-92, or general reference 10-4, pp. 585-595.

$$J = Nq\bar{V} = NqKF$$

we see that the energy input is directly proportional to the mobility of the ions. On the other hand, the energy lost from the positive column is directly proportional to the coefficient of diffusion of the ions, for most of the ions disappear by diffusing to the walls.¹ In equilibrium the energy input must equal the energy loss. But both the coefficient of diffusion and the mobility (with low field strengths) are proportional to the mean free path of the electrons and hence inversely proportional to the pressure. Thus, both the energy input and the energy lost from the positive column are proportional to $1/p$; and since J itself is proportional to $1/p$, we should expect the voltage gradient F to be independent of the pressure as long as the efficiency of ionization remains constant. This is essentially true, as Fig. 10-14 shows. Actually, however, as the pressure is decreased the electrons gain more energy over each free path, and their efficiency for ionization increases. Thus, we should expect a decrease in the voltage gradient with a decrease in pressure. This is also in agreement with Fig. 10-14.

Experiments indicate that, except for its effect on the ion density and the consequent loss of ions by diffusion, the temperature of the positive column has little effect on its characteristics. Actually, it is very difficult to control the temperature of the gas in the positive column by external means. Reference to Eq. (7-10) (Sec. 7-1) indicates that because of the small mass of the electrons their average energy will correspond to an enormously high temperature even when the field strength is only a few volts per meter. Since it is the electrons that carry most of the current in the positive column, even quite large changes in tube temperature will affect the electron energies only slightly. The temperature of the gas in a glow tube is determined principally by the current density in the tube. Experiments indicate that with currents less than 100 ma. in tubes of 1 or 2 cm. diameter the gas temperature probably does not exceed about 200°C.; but, with currents of 1 or 2 amp., the temperature may reach 1200°C. even when the tube diameter is as large as 3 or 4 cm.

¹ Each positive ion carries E , e.v. of energy plus its kinetic energy to the walls, while the energy carried by each free electron is of the order of W_m^* [see Eq. (7-8)].

Most of the energy input to the positive column appears as heat at the walls. Ångström estimated from measurements with a bolometer that in a nitrogen discharge less than 8 per cent of the energy input is radiated in the form of light. Even with efficiencies of this order the glow discharge is several times as efficient a light source as a tungsten-filament lamp. Recently the efficiency of glow discharges as light sources has been increased by coating the inside of the discharge tube with materials that fluoresce and thus convert radiant energy of frequencies outside the visual range into visible light.

Although there are many methods for determining the potential gradient in the positive column, one of the simplest is by the

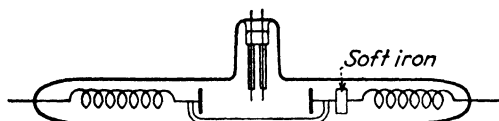


FIG. 10-15.—Tube for determining the positive-column gradient.

use of two probes projecting into the discharge. Figure 10-15 shows a typical arrangement in which the two electrodes are rigidly joined by means of an insulator and can be moved relative to the two probes by a magnet acting on a piece of soft iron attached to the ensemble. Since it is the difference of potential between two probes close together that is measured, many of the errors that occur in the use of a single probe would cancel. However, the use of a single probe has the advantage that by a suitable interpretation of the results obtained not only the potential gradient but the potential of the space and the electron temperature (*i.e.*, r.m.s. velocity of the electrons) and the ion density can also be determined.

10-5. The Striated Positive Column.—In not all cases are the positive column and its gradient uniform. Sometimes the column is broken up into regularly repeated bands or striations of alternate light and darkness. In such a positive column, usually spoken of as striated, the distance between successive striations depends on the pressure of the gas, the current density, the diameter of the discharge tube, and the nature of the gas. Striations are to be seen only within somewhat narrow limits of pressure and current. They are always convex toward the cathode, the surface of the striation being a surface of equal

potential Electric charges on the walls will, of course, affect the shape of the striations. The striations may not always be steady but may wander about in the tube. When this wandering becomes very rapid, it may give an appearance of uniformity to the positive column that does not actually exist. Viewing the discharge in a rotating mirror or through a stroboscope reveals the rapidly moving striations. One remarkable feature about striations is that they do not seem to be produced readily in very pure gases. Thus, in pure nitrogen, they have never been observed; but, on the other hand, by way of exception, they have never been eliminated from even the purest hydrogen.

Figure 10-16 shows the variation in electric intensity and potential throughout a striated discharge. The dotted curve represents the intensity of the light along the discharge. It will be noticed that bright parts of striations are accompanied by maximum electric intensity and dark regions by minimum electric intensity. In many cases the variation of electric intensity is much more marked than is shown in Fig. 10-16. Often the electric intensity approaches zero and in some cases, even, is slightly negative at the cathode side of the striation.

Since the electric intensity is the gradient of the space potential, Fig. 10-16 shows that the space potential need not fluctuate a great deal in order to produce large changes in the electric intensity.

Because the voltage drop between striations is usually close to the ionization potential or first critical potential of the gas, attempts have been made to explain the striations as a resonance phenomenon caused by large numbers of electrons moving with approximately equal energies. The argument is essentially as follows: The various dark and light regions correspond to repetitions of the Faraday dark space and the head of the positive column. Since the voltage drop across a dark space is approximately equal to the first critical potential of the gas, the electrons must all start with approximately zero energy at the cathode edge of the dark space and gain energy uniformly from the electric field. When they have an energy about equal to the first critical

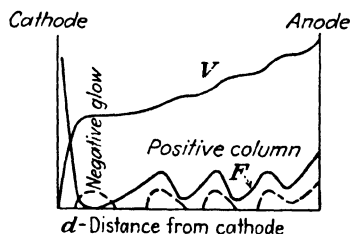


FIG. 10-16.—Variation of potential and electric intensity in a striated positive column.

potential, they can begin to produce excitation. This is where the glow starts. A little later, they have enough energy to produce ionization; but, in ionizing, they lose practically all their energy and cannot then even produce excitation. They must then start over again and gain energy from the electric field. Thus the whole cycle is repeated. Support for this theory is found in the fact that throughout the gaps between striations there is a negative space charge, whereas in the glows the space charge is practically zero and may even be slightly positive on the anode side of the glows.

The absence of striations from discharges in very pure gases has been ascribed to the presence of large numbers of metastable atoms which disperse the electrons and break up the uniformity of the discharge conditions. Slight amounts of an impurity in the gas will give the metastable atoms an opportunity to lose their energy in exciting or ionizing collisions with the molecules of the impurity without greatly disturbing the discharge. The absence of striations from some glows may also be due to the fact that the positive ions formed in the gas move toward the cathode and tend to wipe out the irregularities produced by the electrons. Actually, stationary striations are the exception rather than the rule in the positive column, and it is only under exceptional circumstances that they can be maintained.

Moving striations cannot be explained so simply. It seems now that they are due to, or closely related to, so-called "plasma"¹ oscillations. Thomson² has shown that potential waves can easily be propagated through a highly ionized region of low space charge, such as the positive column of the glow. We now know, principally through the works of Penning³ and Tonks,⁴ that waves or oscillations of this sort can also *originate* in highly ionized (plasma) regions when energy is continuously supplied. Since the frequency of these oscillations may be as high as 10^8 cycles/sec., it is quite easy to understand why they escaped notice for so long. It is not beyond reason that even the sta-

¹ Any highly ionized region in which the electric intensity and net space charge are small has come to be called a *plasma* (see Sec. 10-6).

² General reference 10-1, pp. 353-358.

³ *Nature*, **118**, 301, 1926; *Physica*, **6**, 241, 1926.

⁴ TONKS, L., and I. LANGMUIR, *Phys. Rev.*, **33**, 195, 980, 1929. TONKS, L., *Phys. Rev.*, **37**, 1459, 1931. GUTTON, H., *Ann. Physik*, **13**, 62, 1930. GUTTON, C., *Ann. Physik*, **14**, 5, 1930.

tionary striations may be due to plasma oscillations of such a character that standing waves are set up in the positive column, the wave length corresponding to the distance between striations.

10-6. Determination of Positive-column Characteristics.—The development of a method for the determination of the characteristics of the electrons and positive ions in the positive column is due primarily to the work of Langmuir and Mott-Smith. In 1923, they developed methods by means of which it was possible to interpret the voltage-current curves of small probes placed in the discharge and to determine from them not only the space potential but also the average energy and average density of the electrons and positive ions.

Probe methods have long been in use for determining the potential distribution in a conducting medium, and in the case of solid or liquid conductors the results obtained have been quite accurate. In those cases (see Sec. 2-6) the primary requisite of a probe measurement, namely, that the introduction of the probe into the conducting medium must not alter the potential distribution, was attained when the probe current was zero. However, attempts to determine the potential distribution throughout the space of a gaseous conductor by the same means failed. Until Langmuir and Mott-Smith in 1923¹ developed the proper procedure, probe measurements could be interpreted as giving only a very approximate value to the potentials in an ionized gas space; *i.e.*, the actual space potential might differ from the potential of a probe carrying zero current by as much as 6 or 8 volts.

It is easy to see now why these measurements failed to give the space potential. When the probe current is zero, it is collecting equal numbers of positive ions and electrons. But because of their small mass, the electrons have a much higher velocity than the positive ions even when their average energies are the same. Thus, they will diffuse much faster and soon charge the probe to such a negative potential that positive ions are drawn in as fast as the electrons. In this condition the probe acts the same as any of the insulating walls (see Sec. 10-4). Although it satisfies the conditions that it must not alter the potential distribution in the discharge, its potential is not the

¹*Gen. Elec. Rev.*, **26**, 731, 1923; **27**, 449, 538, 616, 762, 810, 1924; *J. Franklin Inst.*, **196**, 751, 1923; *Phys. Rev.*, **28**, 727, 1928.

same as the space potential. The method of Langmuir and Mott-Smith makes it possible to use a probe to determine the space potential and in addition the average energy and density of the positive ions and electrons. However, the method can be applied only to highly ionized regions, such as the positive column of a glow or arc discharge, where the average density of electrons and positive ions is approximately equal and exceeds 10^{15} per cu. m. In these so-called plasma regions the velocity

distribution of both the positive ions and the electrons is essentially Maxwellian.

Let us suppose that a small plane electrode is maintained in the positive column of a glow discharge at a potential of 20 or 30 volts negative with respect to the space about it. Because of its potential, electrons will be repelled by the probe, leaving a sheath of positive ions adjacent to the electrode. The sheath thickness will adjust itself so that the positive ions composing it will just balance the effect of the negative potential on the probe. In this condition the discharge will be uninfluenced by its presence, and the current collected

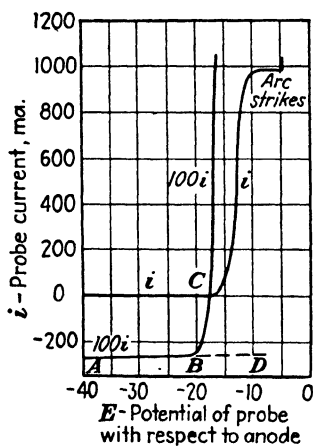


FIG. 10-17.—Positive-column probe characteristic. (I. Langmuir and H. M. Mott-Smith, *Gen. Elec. Rev.*, **27**, 449, 1924.)

by the probe will be just that due to positive ions coming up and being neutralized. The current then will not be greatly affected by the potential of the electrode, for changes in probe potential will result only in changes in sheath thickness, and the area of the sheath (which determines the current collected) will remain essentially constant.

With the probe at a potential higher than the space potential only electrons will be attracted. In this case, there will thus be a negative space-charge sheath, and the current collected will be quite large. At some intermediate potential the current collected will be zero, but this will not determine the space potential. As we saw before, the net current to the walls of the tube was zero; but in order for that to be true the walls assumed a negative potential. So it is with the probe; the current col-

lected may be zero, but the probe potential is negative with respect to the space. This so-called *floating potential* would be identical with the space potential only if the positive and negative carriers were of equal energy, concentration, and mass.

Figure 10-17 represents a typical probe characteristic obtained with a plane electrode in a mercury discharge at a pressure of 0.0252 mm. In the region *AB* the probe is sufficiently negative to repel all electrons; but as the potential is raised, a point *B* is reached where it begins to collect electrons. The floating potential V_f is reached at *C*, but still the probe is at a potential negative with respect to the space. As long as the probe is more negative than the value corresponding to point *B* on the curve, the current is essentially constant and limited by the space charge. Thus, we can use Eq. (6-35) to determine the thickness of the space-charge sheath if we know the current density J and the voltage drop across the space charge E_1 .

$$J = \frac{4\epsilon}{9} \left(\frac{2q_e}{M} \right)^{\frac{1}{2}} \frac{E_1^{\frac{3}{2}}}{d^2} \quad (6-35)$$

In this equation, M is now the mass of positive ions. The magnitude of d is easily verified from visual observations, for the space-charge sheath about an electrode in a plasma is very clearly defined as a dark space about the collector. The positive-ion current to a plane electrode in a positive column is also given by Eq. (4-69) because every positive ion that strikes the sheath boundary is drawn in and eventually strikes the probe. That is,

$$\eta = \frac{N\bar{c}}{4} \quad (4-69)$$

where η is the number of ions striking unit area of the electrode per second, or

$$J = \frac{N\bar{c}}{4} q_e \quad (4-69a)$$

Thus, combining Eqs. (4-69a) and (6-35), we can determine the average velocity or average energy of the positive ions if we know the ion density, the sheath thickness, and the voltage drop.

The number of electrons reaching the probe, when it is maintained at a potential negative with respect to the space, is limited by the fact that the electrons must move against a retarding

potential. Thus, only those electrons reach the probe which have sufficient energy to overcome this retarding potential. The Boltzmann equation [Eq. (6-38)] discussed in Sec. 6-7 can be used to determine the number of electrons reaching the probe; however, since it will yield additional information, we shall apply the relation developed in Sec. 4-15 for the current density of electrons that cross a potential barrier. Equation (4-72) tells us that when the electrons in any region have a Maxwellian distribution of velocities the number of electrons per second of mass m which can cross unit area of a surface against a retarding field, corresponding to an energy W , is given by η .

$$\eta = N \left(\frac{kT}{2\pi m} \right)^{\frac{1}{2}} e^{-\frac{W}{kT}} \quad (4-72)$$

when N is the volume density of electrons and T is the temperature of the electron gas. Since the only assumptions made in deriving Eq. (4-72) were that the electrons obey the gas laws and have a Maxwellian distribution of velocities, this equation can be applied directly to the calculation of the current density of electrons collected by a negative probe in a plasma. Thus,

$$J_e = \eta q_e = N_e q_e \left(\frac{kT_e}{2\pi m} \right)^{\frac{1}{2}} e^{-\frac{W}{kT_e}} \quad (4-73)$$

or, since $W = q_e E_1$,

$$J_e = \eta q_e = N_e q_e \left(\frac{kT_e}{2\pi m} \right)^{\frac{1}{2}} e^{-\frac{q_e E_1}{kT_e}} \quad (4-73a)$$

In this equation E_1 is the voltage drop across the space charge sheath,¹ m and N_e are, respectively, the mass and average density of the electrons in the plasma, and T_e is the temperature corresponding to the average kinetic energy of the electrons in the plasma. This equation can be rewritten in the form

$$\log_e J_e = \log_e B - \frac{q_e}{kT_e} E_1 \quad (10-12)$$

where B is the coefficient of the exponential term in Eq. (4-73a). We can see from this that if $\log J_e$ is plotted against the probe potential a straight line should result whose slope will determine the electron temperature T_e . For the case illustrated in Fig.

¹ E_1 is the difference in potential between the plasma space outside the sheath and the probe.

10-17, i_e (which is, of course, proportional to J_e when a plane probe is used) can be obtained by extending the horizontal portion of the curve AB to D and taking the difference between the ordinates of the actual curve and this straight line. This follows since the total current flowing out of the probe is given by

$$i = i_e + i_p$$

and since i_p —the current due to positive ions coming up to the electrode—constitutes a current in the opposite direction to i_e .

In Fig. 10-18, $\log i_e$ has been plotted as a function of probe potential. It is interesting to note that in this case the points lie on a straight line as far as the point P . The electron temperature T_e , determined by the slope of the curve, in this particular case turns out to be 9200°K . This temperature must not be confused with the gas temperature; for, as we have seen several times before, the electrons can gain a considerable amount of energy in elastic collisions even when the field strength is quite low. In the plasma the electron temperature usually ranges between $10,000$ and $50,000^\circ\text{K}$.

As the probe potential is increased, a point is reached, in the neighborhood of S in Fig. 10-18, where the curve begins to flatten off again. This marks the point where the probe potential is equal to the space potential. Here, the positive-ion space-charge sheath disappears, and there is no longer a retarding force acting on the electrons that are moving toward the electrode. Thus, the current is simply due to the normal motions of the electrons and positive ions. As the potential is raised still further, a negative space-charge sheath is formed at the probe and positive ions are repelled. The total current, however, approaches a practically constant value; for, with the removal of the retarding field, all electrons that enter the space-charge sheath reach the probe. The slight increase in current that occurs with further increase in voltage is due to the fact that the negative space-

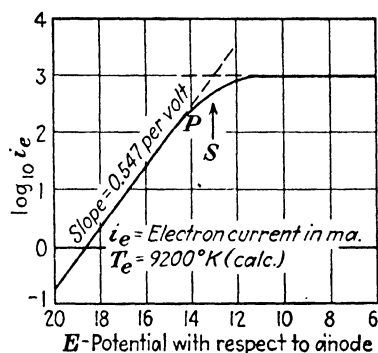


FIG. 10-18.—Electron current collected as a function of probe potential. (*Loc. cit.*)

charge-sheath thickness is many times that of a positive space-charge sheath owing to the longer mean free path and smaller mass of the electrons compared with the positive ions. Because of this, the area of the sheath exposed to the plasma will differ considerably from the area of the probe. Thus, the probe would have to be very large to prevent these edge effects from materially altering the total current collected by the electrode. In practice, the space potential is found quite closely by extending the two straight portions of the curve until they intersect.

At the space potential, E_1 in Eq. (4-73a) is zero and the current density is proportional to the concentration of electrons in the gas. That is, when $E_1 = 0$,

$$J_e = N_e q_e \left(\frac{kT_e}{2\pi m} \right)^{\frac{1}{2}} \quad (10-13)$$

or, since $i_e = J_e A$, where A is the area of the probe,

$$i_e = A N_e q_e \left(\frac{kT_e}{2\pi m} \right)^{\frac{1}{2}} \quad (10-14)$$

This gives us a means of determining the density of electrons and positive ions, N_p , in the plasma, for the two are assumed to be equal. For the case illustrated in Figs. 10-17 and 10-18, N_e or N_p thus comes out to be about 5.83×10^{16} per cu. m., for the probe area is 3.61 sq. cm.

Table 10-3 gives some typical data for the plasma of a low-pressure mercury arc in a 6.2-cm.-diameter tube as determined by probe measurements.

Although the space potential in a plasma region is easily determined by a sharp break in the probe-characteristic curve, characteristics that exhibit the sharp break are the exception rather than the rule. The space potential, however, can usually be located within less than 1 volt, which is more accurate than is possible by any other previous method.

All the discussion in this section so far has been based on the assumption that the velocity distribution of the electrons and positive ions in the plasma is Maxwellian and that modifications would be necessary if the distribution were otherwise. Actually, deviations from the Maxwellian distribution law occur quite often, and probe characteristics do not always show a straight-line dependence of $\log i_e$ on probe potential such as is shown in

Fig. 10-18. In the cases where the velocity distribution does not follow the Maxwellian law, detailed analysis¹ has shown that it is possible, by means of the probe characteristic, to make at least a close guess as to the velocity distribution law that does hold. In these cases, however, an expression for the electron and positive-ion energy in terms of temperature has no meaning, and

TABLE 10-3*

Temperature of liquid Hg, deg. C.	1.4	18.6	38.6
Pressure of Hg vapor, mm. Hg...	0.000214	0.00106	0.00545
Axial field strength, volts/m....	9.32	19.6	31.1
Electron temperature, T_e , deg K.	38,000	27,500	19,900
Electron density at axis of tube (number/cu. m.).....	5.17×10^{16}	9.13×10^{16}	28.6×10^{16}
Number of free electrons in unit length of tube ($2\pi \int N_e r dr$) (number/m.)....	11.1×10^{13}	21.8×10^{13}	45.7×10^{13}
Positive-ion current density to walls of tube, amp./sq. m....	3.56	4.5	5.05
Random electron current in the plasma over the area of the cross section of the tube, i.e., total current borne by elec- trons either way across a cross section of the tube if the dis- tribution in velocity is iso- tropic ($2\pi \int i_e r dr$), amp.....	21.5	35.8	64.0
Drift current in tube, amp.....	5	5	5

* KILLIAN, T. J., *Phys. Rev.*, **35**, 1238, 1930.

the energy must be expressed in terms of the actual distribution law.

It is obvious that there are also other sources of error in probe measurements which can modify the interpretation of the results obtained. For example, the mere presence of the probe will alter the plasma conditions slightly, because it is drawing charges out of the plasma. In most cases the ion density is sufficiently high so that this effect can be neglected. There are other effects, however, the cumulative effect of which may produce considerable error. Some of these are due to inaccuracies in calculating the area of the space-charge sheath, and to the fact that secondary

¹ See general reference 10-2, pp. 337-386, and general reference 10-4, pp. 232-256.

emission may be occurring at the probe surface owing to positive ion and metastable-atom bombardment and the photoelectric effect. In addition, contact potentials at the probe surface or the presence of groups of particles of very high energy will alter the results obtained. The method of analysis, however, is the most powerful that has so far been suggested.

10-7. The Anode in the Glow.—It is well known that the potential drop at the anode of the glow discharge may be anything from a few volts negative to 20 or 30 volts positive or more, depending upon the relative size of the anode and the discharge tube, the kind of gas, and the magnitude of the current. Furthermore the potential drop at the anode may or may not be accompanied by a visible glow different from the glow in the positive column. But because of the difficulties encountered in determining conditions close to the anode, there is still considerable doubt as to what constitutes the true conditions there.

Conditions in the positive column have been fairly well established by means of probe measurements. It is known, for example, that the number of electrons crossing unit area in the positive column in any direction is much larger than is necessary to supply the drift current in the tube (see Table 10-3). Furthermore, although the current carried by the positive ions is small, it still constitutes a very essential part of the tube current. But, at the anode, all the current consists of electrons moving in one direction, namely, into the anode. Any satisfactory explanation of conditions at the anode must thus explain the transition from the positive column, where nearly equal numbers of electrons are moving in all directions, to the anode, where electrons are moving in only one direction. In addition, it must account for the production of enough positive ions near the anode to supply those necessary for the positive column as well as those lost to the walls.

The general approach has been to assume that a space-charge sheath is established at the anode in which these functions are accomplished. Ordinarily, this is a negative space-charge sheath; but, under some conditions, it can be a positive sheath. The negative sheath arises owing to the fact that positive ions are repelled by the positive electrode. The electrons, on the other hand, are accelerated and must produce an appreciable amount of ionization by collision in the sheath in order that enough positive ions may be produced to supply the requirements of the positive

column. The voltage drop across this sheath is thus approximately equal to the ionization potential of the gas. If the anode is very small, the space-charge sheath may have an area considerably larger than the area of the anode. This is necessitated by the fact that in this condition the product of the random electron current density in the positive column and the anode area is less than the current flowing in the external circuit.

When the anode is large and practically fills the discharge tube, it is possible for the anode fall to be actually negative. When this condition exists, the product of the random electron current density in the positive column and the area of the anode is larger than the current in the external circuit and some of the electrons approaching the anode must be turned away. Thus, a positive space charge develops near the anode.

The bright balls of glow that often appear on the anode in a glow discharge are further evidence of the presence of a space-charge sheath. These balls of glow are usually due to local regions of high vapor density caused by the emission of gas from the anode. Gas is not necessarily emitted uniformly from the surface of any electrode; and when gas emission takes place, the current density increases in these regions and causes the glows to appear.

An exceptionally high anode fall of potential has been observed in some cases when the gas contained compounds of the halogens. The explanation for this has been that, since the halogens have a very high affinity for electrons, there are very few free electrons in the gas. Then, in order to produce sufficient positive ions for the positive column the voltage drop across the negative space-charge sheath will have to be large enough so that there is sufficient ionization by collision.

10-8. Stability of Discharges.—The voltage-current characteristic of any gaseous or vacuum discharge is determined by conditions in the discharge region for any given tube. This means that the characteristic is independent of the external circuit and depends only on the balance between the ionizing and deionizing agents active in the tube. This is evident from a consideration of any of the discharges discussed thus far. For example, the glow discharge has a voltage-current characteristic that is concave upward. The voltage drop across the discharge falls rapidly with an increase in current after a self-maintaining discharge has been formed and attains a value, in the normal

glow in a short tube, slightly higher than the minimum sparking potential. Then, if the current is increased until an abnormal discharge is formed, the voltage drop again increases. In the case of the glow discharge, this entire characteristic can be explained on the basis of the ionizing and deionizing agents active in the tube under any given conditions. Similarly, in the cases of the thermionic diode or the phototube, the voltage-

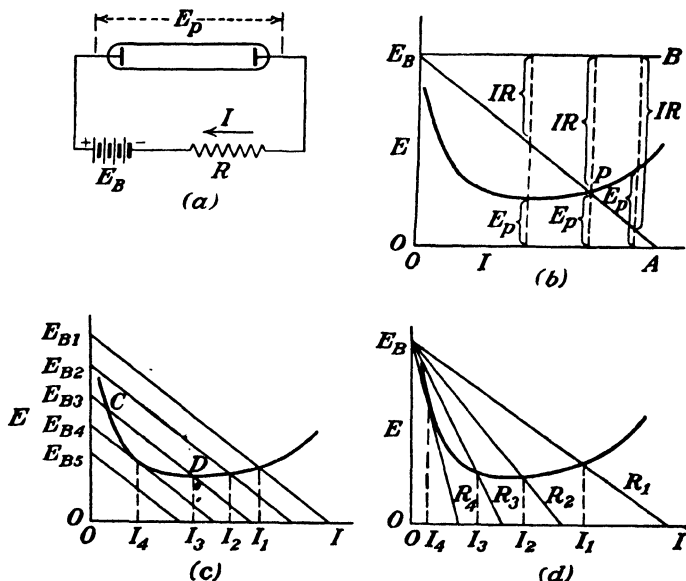


FIG. 10-19.—Effect of circuit voltage and resistance on the stability of a discharge. current characteristic is dependent only on conditions in the tube itself.

The current that flows in an actual test, however, is not independent of the external circuit. This was pointed out in Sec. 6-11; but it is important to examine the situation in more detail at this time, for the external circuit plays such an important role in the operation of many gas discharge devices. In any circuit, we must have Kirchhoff's laws satisfied. Thus, for example, in the simple series circuit shown in Fig. 10-19a, we must have at all times

$$E_B = E_p + IR \quad (10-15)$$

In other words, if a glow discharge is formed, the current in the external circuit (and in the tube) must adjust itself to such a

value that the sum of the voltage drops across the glow and the series resistance is just equal to the impressed voltage. If the external resistance is small, a very large current may be necessary to satisfy this condition.

In a particular case, we might have the situation illustrated in Fig. 10-19b. Suppose that the curve in this figure is the volt-ampere characteristic of the discharge tube under consideration. Suppose, also, that the external resistance in the circuit is such that the ordinates to the line $E_B A$ represent $(E_B - IR)$. If the battery voltage E_B is higher than the sparking potential in the tube, then as soon as the switch is closed in the external circuit the current will increase to a value corresponding to the intersection of the two curves at point P , for at this point the sum of the voltage drops in the tube and in the external resistance is equal to the supply voltage. No other value of current is possible; for if the current tends to increase to a value higher than that given by the intersection of the two curves, $(E_p + IR)$ will be larger than E_B and the current will immediately decrease again. Similarly, if the current tends to decrease from the point of equilibrium, $(E_B + IR)$ will be less than the circuit voltage and the current will immediately start to increase. Thus, the point of intersection is the stable operating point, and it is impossible to have any other value of current when the circuit voltage and circuit resistance remain constant.

The effect of changes in battery voltage are illustrated in Fig. 10-19c. As long as the circuit resistance is maintained constant, changes in circuit voltage simply have the effect of moving the resistance line up or down. For the case illustrated, successively lower values of current are obtained in the circuit as the supply voltage is reduced from E_{B1} to E_{B4} . The relationship between the circuit current and the supply voltage is not a linear one, however; for the glow characteristic is, of course, not a straight line. Furthermore, it is impossible to maintain a discharge with a current smaller than I_4 if E_B is the only variable; for when E_B is less than E_{B4} , we find that, for every value of current, $(E_p + IR)$ is larger than the supply voltage and that there thus is no stable value of current.

The question quite logically arises as to why we can have only one stable value for the current when we have a battery voltage such as E_{B3} and the resistance line cuts the glow characteristic

in two points. There are certainly two points where Eq. (10-15) is satisfied, but the point D is the only stable operating point. It is easy to see that this is true by supposing that it is possible to have a glow corresponding to the intersection of the two curves at C . In this condition, if the current should decrease ever so slightly, we should find that $(E_p + IR)$ would be greater than E_B and that the current would continue to decrease until the discharge went out. On the other hand, if the current should increase slightly, $(E_p + IR)$ would be less than E_B and the current would continue to increase until Eq. (10-15) is satisfied at the stable point D . Further fluctuations in current would, as explained above, always bring the current back to I_3 .

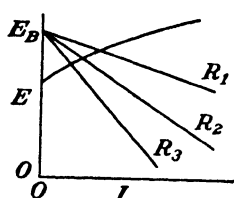


FIG. 10-20.—Stability in a discharge with a rising volt-ampere characteristic.

The effect of variations in circuit resistance is shown in Fig. 10-19*d*. When the circuit voltage remains constant, an increase in the external resistance reduces the current flowing; but again the relation between R and I is not linear, for the glow characteristic is not a straight line. As the resistance increases from R_1 to R_4 , the slope of the resistance line increases and the current flowing decreases from I_1 to I_4 . There is a limit to the minimum value of current that can be obtained in this way, just as there is a limit to the current variation that can be obtained by changing the circuit voltage; for when the resistance line becomes tangent to the glow characteristic, the minimum value of current has been attained. Practically the entire glow characteristic can be traced out, however, if a high enough circuit voltage and resistance are used.

If conditions in the discharge tube are such that the sparking condition corresponds to a point on the left of the minimum of the sparking-potential curve, the volt-ampere characteristic for the glow will be as indicated in Fig. 10-20. With any discharge having a rising volt-ampere curve such as this, a stable operating point can be obtained with any circuit resistance and any circuit voltage as long as the resistance line intersects the tube characteristic.

10-9. Applications of the Glow Discharge. *a. Glow-tube Oscillator.*¹—The fact that the glow discharge has a falling volt-

¹ For further discussion of the gaseous-discharge oscillator, see Sec. 11-6.

ampere characteristic suggests its possible use to maintain an oscillating circuit. In general, the types of oscillation produced by a glow discharge can be grouped into two classes. In one, the glow is more nearly described by saying that it is intermittent; in the other, there are definite pulsations in the glow current and voltage, although the glow never goes out. In neither case does the current through the tube oscillate, although the current in a parallel circuit may.

Figure 10-21 represents one type of circuit. When the switch S is closed, the potential starts to rise at the terminals of the glow tube. The rate at which the potential rises is limited by the resistances R and r in-series with the condenser C . When the condenser has been charged to a potential equal to the sparking potential of the tube T , a breakdown occurs. The voltage across T immediately falls, and the condenser begins to discharge through the glow. If r is not large, the current flowing in T will be large and the voltage will be correspondingly low, on the assumption, of course, that the electrodes in T are sufficiently large so that the discharge never reaches the abnormal state. In a very short time, the condenser will discharge to a voltage equal to the glow voltage, and the current will decrease. The glow-tube voltage will then tend to increase; however, it cannot increase faster than the voltage rises on C , and that rate of rise is limited by the value of R . Thus, when the current through T decreases, the glow voltage will not be able to increase as it should along the volt-ampere characteristic. With an insufficient voltage, the glow goes out, and the voltage across C again builds up to the sparking potential, when another glow is formed and the cycle is repeated. During the time that C is being charged, the potential across r is in one direction; while C is discharging, it is in the opposite direction. Thus an alternating potential is made available. If the resistance r is replaced by an inductance or the primary of a transformer, extinction of the glow is facilitated; for then the current in the condenser branch is made to persist, by the inductance, after the condenser voltage has fallen to the stable glow voltage. The condenser discharges then to a still lower voltage; and when the current through the inductance

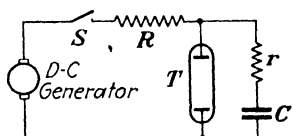


FIG. 10-21.—Intermittent glow oscillator.

reverses, all current may be diverted from the glow which then is immediately extinguished. The maintenance of an intermittent discharge of this type depends to a large extent on the choice of proper values for R and C .

The other type of oscillation is obtained with a glow or arc that is never extinguished, although the current may pulsate rapidly or oscillate about an average value. The circuit for this is represented in Fig. 10-22. Oscillations in this circuit can be obtained only if the inductance of the supply circuit L' is large enough to prevent rapid fluctuations in current. Suppose that R has been adjusted so that the glow discharge corresponds to some point A on the falling portion of the volt-ampere curve (see

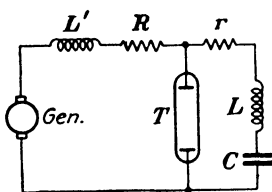


FIG. 10-22.—Continuous glow oscillator.

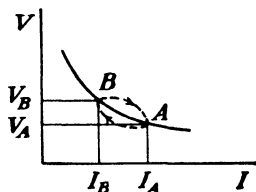


FIG. 10-23.—Volt-ampere characteristic of a glow oscillator.

Fig. 10-23) and that L and C have been adjusted so that oscillations, once started, will be maintained. Let us see how it is possible to have sustained oscillations in the circuit of L , C , and r . If the current in T should suddenly decrease, owing initially, perhaps, to a sudden change in R , the voltage of the glow will increase along the volt-ampere characteristic to B . At the same time the current in r , L , and C will increase by the amount of decrease in T , for the current through L' and R is constant. Diverting some of the current of T into C means that the voltage of C will increase. It will increase to V_B . But then, because there is inductance in series with C and the current cannot stop immediately, the voltage will rise still higher, and finally the current in C will come to zero, with the voltage of the condenser greater than V_B . The condenser will then start to discharge through T . But, in the meantime, the current that has been flowing into C has been turned back into T ; and with the increase of glow current from I_B to I_A the voltage falls to V_A , and C discharges rapidly. Again, when the voltage of C has fallen to V_A , the current cannot cease but continues, owing to L . C dis-

charges to a voltage below V_A . Now, again, current is diverted from T to recharge C ; and again the tube voltage rises, and the cycle is repeated. Because a definite time is required for equilibrium to be established in the glow or arc, the tube voltage does not follow the static characteristic exactly, but changes as shown by the dotted curve. The finite hysteresis loop shows that there is a quadrature component in the voltage change, with the voltage leading the current. Thus, the glow or arc behaves as if it possessed inductance. For this reason the frequency of oscillations of this type is less than the natural frequency of the circuit containing L and C . If one attempts to attain too high frequencies with this circuit, a rising dynamic volt-ampere characteristic may be obtained and oscillations will not be maintained. In most cases, frequencies of only a few thousand cycles are possible. In hydrogen, however, an equilibrium of the ionization is attained much more rapidly, and it is possible to follow the static characteristic closely with frequencies much higher than is possible with other gases.

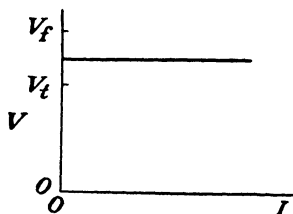


FIG. 10-24.—Volt-ampere characteristic of ideal lightning arrester.

b. Lightning Arresters.—The volt-ampere characteristic of an ideal lightning arrester is a horizontal line, as shown in Fig. 10-24. An ideal lightning arrester should pass no current until a critical potential is reached, and then it should pass any amount of current with no change in voltage. With such an arrester on a piece of apparatus, only a slight rise in voltage would occur from a lightning disturbance. On a transmission line, for example, the peak circuit voltage might be represented by V_i in Fig. 10-24, and the voltage at which the insulators would flashover might be represented by V_f . The lightning arrester then, if it is to function properly, must not allow the voltage to rise above V_f ; for then the insulators would flashover and produce a short circuit on the line. Furthermore, it must not allow the circuit voltage to drop below V_i , for this might produce instability in the line. But it must carry away to ground any charges that build up on the line owing to lightning disturbances in the neighborhood of the transmission line. The only way it can do this is to have a nearly horizontal volt-ampere characteristic.

Our study of glow discharges has shown that it is possible to obtain this type of characteristic with a normal glow discharge when the electrode spacing and the pressure are such that the sparking potential has its minimum value. Under these conditions, there is no drop in voltage when the glow starts; and until the cathode is completely covered by the negative glow, the voltage drop across the discharge remains constant. All effects of a positive column in increasing the voltage drop can be eliminated by making the electrode spacing small enough to prevent the formation of a positive column. The fact that the normal cathode fall in a glow discharge is only about 250 or 350 volts and much lower than the operating voltage of most transmission-line equipment is not a disadvantage, for the operating voltage of the lightning arrester can be built up to any desired value simply by using enough discharges in series. Equation (10-2) tells us that the normal current density in a glow discharge is proportional to the square of the pressure. Thus, by using high pressure (atmospheric), large currents can be passed by the normal glow even with relatively small electrodes. The use of atmospheric pressure, however, means that the electrode spacing will have to be about 0.001 cm. if the minimum sparking potential is to be obtained. Because it is practically impossible to have this electrode spacing without having bridging contacts somewhere that will short-circuit the electrodes, the problem has finally been solved by using electrodes of relatively high resistivity (graphite or carbon mixtures usually) in actual contact.¹ With this arrangement, leakage currents are kept at a minimum because of the high resistance of the contact points. Furthermore, somewhere in the neighborhood of the contact points the spacing corresponds to the minimum on the sparking-potential curve. After sparking occurs, the voltage between the electrodes corresponds to the normal glow voltage until the cathode is completely covered by the glow. The total voltage, however, as measured at the terminals of the arrester, should rise slightly with current, owing to the resistance drop in the electrode disks; and thus an exactly horizontal characteristic cannot be expected.

Cathode-ray oscillograms of the volt-ampere characteristic of arresters built along these lines appeared as in Fig. 10-25 and showed that the predicted variation of voltage with current (the

¹ SLEPIAN, J., *Trans. A.I.E.E.*, **45**, 169, 1926.

dotted curve) was approached only as the current decreased from its maximum value. The fact that on the application of a surge voltage the voltage between the electrodes of the arrester rose higher than the predicted curve could be explained as due to the definite time required for the normal glow to spread over the cathode surface, from the points at which the discharge started. Thus, actually, an abnormal glow discharge developed as the glow spread out over the surface of the electrodes. However, the voltage could not rise higher than the sparking potential for the greatest gap between the electrodes, for then breakdown would have occurred everywhere between the electrodes. When the current decreased, the predicted curve was followed closely; for the contraction of the glow involves only deionization, and that can occur without a change in voltage.

Actually, with the electrodes in contact, it was found that the glow voltage did not rise more than about 100 volts per gap. This is not serious, for the range in voltage between the peak line voltage and the insulator flashover voltage is usually wide enough to accommodate this fluctuation in the arrester voltage. In practice, a spark gap is usually used in series with the arrester to interrupt the leakage current which would otherwise flow from the line through the electrode contacts.

c. The Vacuum Switch.—Since the sparking potential of any gap rises rapidly as the pressure is reduced below a critical value, very high voltages are necessary to cause sparking in high vacuum. Because of this fact, vacuum switches are very useful for high-speed a.c. switching. These switches can be made very compact, for with the contacts separated only a fraction of an inch the gap can withstand a very high voltage. In actual operation the circuit is not opened immediately on separating the contacts because an arc always forms which persists for the duration of the half cycle. As a matter of fact, the presence of an arc is very desirable; for if the circuit were opened immediately on the separation of the contacts, high voltages could be developed in any inductances that happened to be in the circuit. By allowing the current to continue until the end of a half cycle, the inductances have a chance to dissipate their stored energy

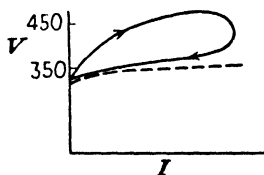


FIG. 10-25.—Volt-ampere curve of lightning arrester.

and the circuit can then be opened without the development of surges.

As the contacts separate, a cathode spot forms on the negative electrode and keeps the arc supplied with vapor by continually evaporating some of the negative electrode material. In other words, by the evaporation of the electrode material a pressure high enough for a low-voltage discharge is maintained. Now as the current passes through zero, the stream of vapor does not stop immediately owing to the thermal lag of the electrode material but continues to stream from the point on the electrode where the cathode spot has been. The potential has reversed, however; and if the voltage rises rapidly enough, a glow discharge will form between the two electrodes (see Fig. 10-26), for the vapor from



FIG. 10-26.

the old cathode spreads out and raises the vapor density in the region between the two electrodes. Because of the relatively low pressure, the glow voltage is high and the current density is small. Furthermore, as the old cathode cools off and emits less and less vapor, the vapor density in the region between the electrodes becomes smaller and smaller and the voltage necessary to maintain a given current increases continuously, as shown in Figs. 10-7 and 10-8. In other words, as the vapor density decreases, the voltage-current characteristic shifts rapidly toward lower currents and higher voltages. Thus, unless a new cathode spot is formed to furnish a new source of vapor, the discharge ceases before the next reversal in polarity occurs.

This type of switch is obviously not applicable to d.c. circuits, for in such a circuit there is no reason for the arc that forms on separating the electrodes ever to be extinguished, unless, of course, the contacts are separated a considerable distance, the current is less than about 0.1 amp. (1 or 2 amp. in low-voltage circuits), or special oscillating circuits are used to extinguish the arc. These limitations are not present when alternating current is used, for the circuit is opened as soon as the potential reverses and the current passes through zero.

Problems

1. In a certain hydrogen glow discharge between aluminum electrodes the normal cathode dark-space thickness is 1.45 mm. when the pressure is

5 mm. Hg. Estimate the average number of collisions made by a positive ion and an electron in crossing the dark space. Assume a gas temperature of 0°C .

2. Using the data of Fig. 8-10 and Table 10-2, estimate the average number of ionizing collisions made by an electron in crossing the cathode dark space of the discharge described in Prob. 1.

3. What would be the normal cathode dark-space thickness of the discharge described in Prob. 1, if the pressure were (a) 0.1 mm. Hg? (b) Atmospheric?

4. If the normal current density in the discharge described in Prob. 1 is 1.88 ma./sq. cm., what would be the normal current density at (a) 0.1 mm. Hg? (b) Atmospheric pressure?

5. At a pressure of 2.5 mm. Hg, a certain abnormal glow discharge has the following characteristics: $V_a = 232$ volts when $J_a = 1$ ma./sq. cm. and $V_a = 244$ volts when $J_a = 100$ ma./sq. cm. What pressure would be necessary in the tube to cause a voltage increase of 50 volts in the same current range?

6. The total current collected by a plane probe, 1.9 cm. on a side, lying flush with the tube wall in the positive column of a mercury discharge, varied in the following manner with the applied potential, all potentials being measured with respect to the cathode:

E , volts	I , ma.	E , volts	I , ma.
-100	-0.965	23	5.75
- 50	-0.815	24	9.1
- 30	-0.755	25	14.27
- 20	-0.725	26	21.9
- 10	-0.695	27	33.8
0	-0.665	28	51.7
10	-0.609	29	79.2
15	-0.403	30	121.1
17	-0.108	31	185
18	+0.160	32	265
19	+0.567	33	286
20	+1.187	34	288
21	+2.13	35	290
22	+3.57	36	292

Determine: (a) the space potential in the plasma at the probe, (b) the average kinetic energy of the electrons in the plasma in degrees Kelvin and in equivalent electron volts, (c) the average concentration of the electrons, and (d) the average concentration of the positive ions.

7. Using the data of the preceding problem, compute the voltage drop across a positive-ion space-charge sheath at the probe when the total positive-ion current collected is 1.1 ma. and the sheath thickness 0.29 cm.

8. Estimate the value of series resistance necessary to give a current of 4 ma. in a glow tube having the voltage-current characteristic shown in Fig. 10-3, when the circuit voltage is 1,600 volts.

9. Estimate the smallest value of current that could be obtained in the glow discharge whose characteristic is shown in Fig. 10-3, when it is connected through a resistance to a 1,200-volt circuit. What value of series resistance is necessary to give this current?

10. Estimate the minimum value of circuit voltage that will give a stable discharge in the glow tube whose characteristic is given in Fig. 10-3, when the series resistance in the circuit is 300,000 ohms. What is the minimum glow current that can be obtained under these conditions?

General References

- 10-1. THOMSON, J. J., and G. P. THOMSON: "Conduction of Electricity through Gases," Vol. II, Chap. VIII, Cambridge University Press, London, 1933.
- 10-2. DARROW, K. K.: "Electrical Phenomena in Gases," Chaps. XI and XII, The Williams & Wilkins Company, Baltimore, 1932.
- 10-3. v. ENGEL, A., and M. STEENBECK: "Elektrische Gasentladungen," Vol. II, Verlag Julius Springer, Berlin, 1934.
- 10-4. LOEB, L. B.: "Fundamental Processes of Electrical Discharge in Gases," Chap. XI, Part A, John Wiley & Sons, New York, 1939.

CHAPTER XI

THE ARC DISCHARGE¹

11-1. Introduction.—Studies of the electric arc date back to 1808–1810 when Davy discovered that an exceedingly bright discharge could be formed by bringing together and then separating two pieces of hard carbon which had been connected to the terminals of a battery of about 100 volts. From its shape when the carbons were placed horizontally end to end, he called it an “arc discharge.” Since that time, however, many discharges have been studied that have properties similar to the carbon arc, and the term is now applied to a wide variety of discharges. We speak of arcs at atmospheric pressure and low-pressure arcs, hot-cathode and cold-cathode arcs, self-maintaining and non-self-maintaining arcs. Yet in two respects these are all alike. A certain amount of gas or vapor is always necessary to maintain the discharge, and the potential fall at the cathode of the discharge is always less than the minimum sparking potential for the gas and is usually not more than 10 or 20 volts. This last characteristic furnishes the principal distinction between the glow discharge and the arc discharge.

Our study of the glow discharge has shown us that, since a glow discharge is maintained entirely by ionization by collision and secondary-electron emission from the cathode, a voltage drop at the cathode approximately equal to the minimum sparking potential is necessary in order to have a self-maintaining discharge. This means that a voltage drop of between 200 and 300 volts is usually necessary at the cathode. Since an arc can be maintained with only a fraction of this voltage drop, some new ionizing agent not active in the glow must be aiding the maintenance of the arc discharge. One suggestion is that the cathode becomes hot enough to emit electrons thermionically and that most of the current is carried by these electrons. This condition is sufficient to explain the existence of low-voltage carbon arcs, for example; but it is not sufficient to explain the existence of

¹ General references for Chap. XI will be found on p. 459.

low-voltage arcs between metals such as mercury, aluminum, or copper which boil at temperatures too low to account for an appreciable amount of thermionic emission. In these cases, some other mechanism must be active, and it has been suggested that a dense sheath of positive ions forms so close to the cathode surface that the resultant electric field at the cathode is large enough to provide electrons for the arc by high-field emission. These theories will be discussed in detail in Sec. 11-4. In addition to these two types of self-maintaining arc, there are, of course, the non-self-maintaining arcs such as in the phanotron, thyatron, and tungal rectifier tubes which depend upon an external source of energy to maintain a thermionic cathode for the discharge. In respects other than in the phenomena at the cathode, there is little to distinguish an arc from a glow discharge except that the current density in the positive column of the arc is usually, but not necessarily, greater than in the glow discharge.

The usual method of igniting an arc is to bring the two electrodes together and then draw them apart. As they separate, an arc will form. Except in the case of an arc whose cathode is heated by external means, it is not possible, in general, to establish an arc simply by raising the potential between two electrodes in a gas unless the voltage is raised to the sparking potential.¹ As we saw in Chap. IX, a voltage equal to the sparking potential is necessary for the establishment of a self-maintaining discharge. However, when the sparking potential is reached, the discharge can take the form of a corona, a glow, or an arc discharge. Both the corona and the glow discharges will change over to arc discharges if the current is increased to a large enough value. The transition from corona to arc occurs suddenly at a fairly definite value of electrode voltage which depends upon the electrode configuration (see Sec. 9-6). The transition, however, represents an unstable condition and ordinarily cannot be traced out completely point by point.

Usually, the transition from glow to arc discharge takes place suddenly and discontinuously at no definite value of voltage; but if the electrodes are tungsten or carbon and a circuit of very high resistance is used, it is possible to observe a continuous change. Figure 11-1 shows such a continuous transition from the glow type of discharge to the arc type. This

¹ Exceptions to this occur under special conditions (see Sec. 11-11).

figure refers to a discharge between small (3 mm. diameter) tungsten spheres in hydrogen at 40 cm. Hg pressure. Let us trace out the various stages of the transition. As the potential is raised between the two electrodes, the sparking potential is ultimately reached and a glow discharge forms. If the circuit resistance is high, a normal glow discharge is formed and the voltage between electrodes falls to a value at *A* approximately equal to the minimum sparking potential. After the cathode is completely covered by the glow discharge, any further attempt to increase the discharge current by increasing the supply voltage results in an increased voltage drop between the electrodes. That is, the condition of an abnormal glow is reached. When the discharge corresponds to a point to the left of *A*, an increase in current is accompanied by a decrease in voltage so that the net heat input to the cathode remains essentially constant or increases only slightly. In the region of the abnormal glow between *A* and *B*, an increase in current is accompanied by an increase in the cathode fall of potential. Thus the heat input into the cathode is increased with the current, and the cathode temperature begins to increase (see curve marked *T*). Up to the point *A* the cathode temperature is not high enough to account for an appreciable amount of thermionic emission. But as the heat input increases and the cathode temperature rises owing to the abnormal glow, thermionic emission starts. At *B*, the temperature of the cathode has become high enough to make thermionic emission an important, although still not the primary, source of ionization. Because of the exponential dependence of thermionic emission on electrode temperature, only a slight amount of additional heat is now required to increase the emission considerably. Thus, it is possible for the voltage between electrodes to decrease as the current is increased. Between *B* and *C*, we have a region that in low-resistance circuits

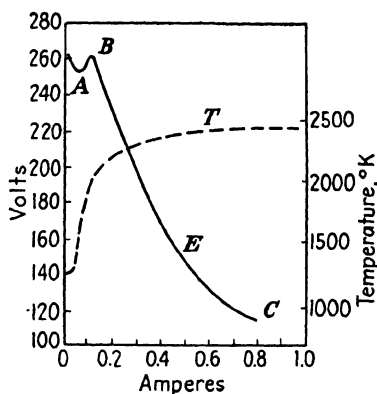


FIG. 11-1.—Glow-to-arc transition with tungsten electrodes. (R. Seeliger, Wein-Harms, "Handbuch der Experimental Physik," Vol. 13, Part 3, p. 612.)

would be unstable but in high-resistance circuits can be traced completely. Further increase in current means more heat input to the cathode, and consequently a higher cathode temperature and more thermionic emission. Thus, the discharge can exist at lower and lower voltages. The minimum voltage that can be obtained in this way needs to be only large enough to allow for the production of some positive ions by ionization by collision. Since it is the bombardment of the cathode by positive ions that keeps it hot enough for thermionic emission, a certain number of positive ions must always be present; in addition, if the positive ions were absent, an electronic space charge would develop which would limit the current. In the region between

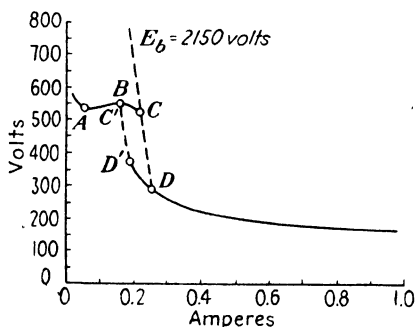


FIG. 11-2.—Glow-to-arc transition with tungsten electrodes. (*Loc. cit.*, p. 613.)

B and *C* an arc discharge is said to exist in which most of the current is carried by electrons generated thermionically. The carbon arc and the type S-1 sun lamp are examples of this type of arc.

A continuous transition such as is indicated in Fig. 11-1 can be obtained only under special conditions. Usually, the transition from glow to arc is discontinuous and erratic, occurring at one value of current in one test and another value of current in the next test. In many cases the glow and arc characteristics actually overlap so that for some values of current either a glow or an arc is possible. Figure 11-2 illustrates this type of volt-ampere characteristic. It refers to a discharge between tungsten electrodes in nitrogen at a pressure of 65 mm. Hg, the electrode separation being 14 mm. At low values of current, this characteristic is similar to that in Fig. 11-1. At *A* a normal glow discharge has been formed; and between *A* and *B* an abnormal glow

is present, and the glow voltage increases with the current. At the same time the cathode temperature starts to rise and may actually reach a high enough value so that with some thermionic emission the voltage between electrodes actually starts to decrease; but finally a point *C* is reached where the form of the discharge suddenly changes, and at the same time the voltage falls to a much lower value. This transition is indicated by the dotted line between *C* and *D*. The total current will increase in this process only enough to allow the difference between the arc voltage and the supply voltage to be taken up in the external resistance; but the cathode current density increases by a factor of 50 or more. The point at which this sudden change-over from glow to arc occurs is not always a constant. When the current is increased suddenly, it may be possible to remain on the glow characteristic even with a current much higher than indicated by point *C*. On the other hand, if the glow is maintained at a point between *B* and *C* a sufficiently long time, the transition is still apt to occur. Similarly, when the discharge is in the arc form and the current is reduced by decreasing the circuit voltage, the arc will change back to a glow (or go out if the supply voltage is too low) at some value of current corresponding to *D'*, for example. The point *D'* is somewhat variable, just as the point *C* is, and depends on how rapidly the arc current is varied. It is always found, however, that the transition from *C* to *D* corresponds to a larger current than the reverse transition from *D'* to *C'*; and, in high-resistance circuits, *C* may actually correspond to a larger current than *D'*. Thus, it is possible, in a limited range of current, to have either a glow discharge or an arc, depending upon the previous history of the discharge.

The continuous transition such as is indicated in Fig. 11-1 is never obtained with electrodes such as mercury, copper, and gold which melt at temperatures low in comparison with the melting point of tungsten or carbon. Experiments have shown that when the peak of the abnormal-glow characteristic (point *B*, Figs. 11-1 and 11-2) is reached the cathode temperature has reached a temperature of at least 2000°C. In the case of low-melting-point materials, cathode temperatures of this order are not possible; and in these cases the glow always changes suddenly to an arc, frequently even in the region of the normal glow discharge. Figure 11-3 illustrates a situation of this sort. The character-

istic was obtained by using gold electrodes in air at a spacing of 3.5 mm. In this case, no abnormal discharge was obtained before the glow changed over to an arc. With electrodes of low melting point the sudden change in cathode current density at the transition is much more pronounced than with electrodes of carbon or tungsten. At the transition the current density changes from a fraction of an ampere per square centimeter to thousands of amperes per square centimeter.

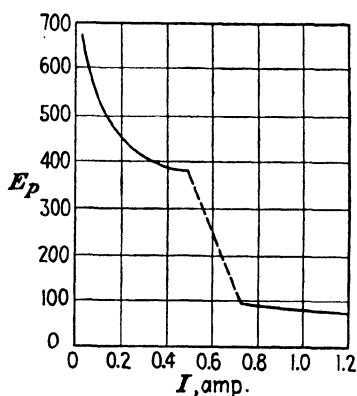


FIG. 11-3.—Glow-to-arc transition with gold electrodes. (*H. E. Ives, J. Frank. Inst.*, **198**, 456, 1924.)

As was mentioned previously, it is possible to initiate an arc discharge without raising the electrode voltage to the sparking potential when an externally heated cathode is used. In a case of this sort, current starts to flow between the electrodes as soon as the electrode voltage exceeds zero. The current flowing, however, is at first purely electronic; but as soon as the electrode voltage rises a few volts, some of the higher speed electrons are able to acquire enough energy to produce ionization and the current increases at a much faster rate than is possible in a high-vacuum tube. Furthermore, with the onset of ionization by collision, positive ions are produced which neutralize the negative space charge initially present and allow the current to increase still more rapidly. By the time the tube voltage reaches 15 or 20 volts, the electrons are producing so much ionization by collision that a stable arc discharge is formed and the current is no longer limited by conditions in the tube. Further increases in circuit voltage cause an increase in current, but the potential drop in the tube remains practically constant. This type of arc is encountered in the phanotron rectifier and in the thyatron.

With a proper kind of filamentary cathode (a tungsten filament, for example), it is sometimes possible to open the filament heating circuit and still maintain the arc discharge after it has once been initiated by means of the filament. In these cases the cathode is maintained hot enough by positive-ion bombardment

so that the discharge can persist. This principle is used to maintain the arc in some types of gaseous-discharge lamp.

11-2. Total Arc Characteristics.—Figure 11-4 is a typical characteristic showing the distribution of potential in a 6-amp. copper arc in air. One is immediately struck by the fact that the curve seems to be made up of three distinct parts, a rather sharp fall in potential at the anode of 7 or 8 volts, a practically uniform decrease in potential in the positive column, and another sharp fall in potential at the cathode of about 20 volts. The anode fall can be explained as due to a space-charge sheath in exactly the same way as is the anode fall in the glow discharge (see Sec. 10-7). It might be added, however, that for high-pressure discharges this space-charge sheath is usually negative. Experimental verification for this is found in the fact that when certain salts and oxides are sprinkled on the anode of a high-pressure arc the anode fall is reduced. Since it is known that these salts and oxides emit positive ions rather easily when heated, the decrease in anode fall is easily explained as the reduction in the space charge near the anode by the emission of the positive ions. Conversely, it has been found that increasing the negative space charge at the anode has the effect of increasing the anode fall. This last effect is particularly noticeable in the case of a carbon arc where the temperature of the anode is about 4000°K . and is emitting a large number of electrons thermionically.

The sharp anode fall of potential apparently has a much more important bearing on the characteristics of high-pressure arcs in air than one ordinarily associates with the positive electrode of a discharge. The falling volt-ampere curve so characteristic of arcs in air at atmospheric pressure can be explained as due in part, at least, to the temperature that the anode assumes. At pressures of 1 atm. or more the space-charge sheath at the anode is so thin that practically all the heat generated there owing to current flowing through the potential drop goes to heating the anode; and unless special provisions are made for cooling, its

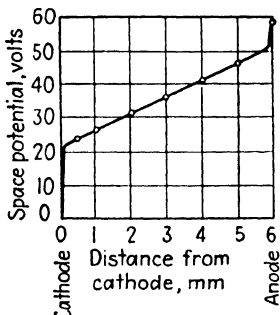


FIG. 11-4.—Voltage distribution in a copper arc in air. (W. B. Nottingham, *J. Frank. Inst.*, **208**, 192, 1929.)

temperature easily rises to the boiling point. The vapor introduced into the arc stream from the anode can then materially affect the characteristics of a short arc.

The cathode fall of potential will be discussed in a later section; and so, for the present, nothing more need be said about it except to point out that it, too, arises because of the presence of a space-charge sheath. One other point should be emphasized, however. Although the anode fall may vary from a few volts to 50 or 100 volts in particular cases, the cathode fall is never more, or for that matter less, than 10 or 20 volts.¹ Furthermore, it is the vapor from the cathode rather than the anode that frequently determines the characteristics of *low-pressure*, or so-called *vacuum*, arcs.

The only difference between the positive column of a high-pressure arc and the positive column of a glow is that usually the current density is higher in the arc than in the glow.² In both the arc and the glow the positive column is a *plasma* region having approximately equal densities of electrons and positive ions and having a relatively small potential gradient. Although the average electric intensity in the positive column of an arc is greater than in a glow discharge (for the example illustrated in Fig. 11-4, it is 4,700 volts/m.) the ratio of the positive-column gradient to the pressure is considerably less in an arc than in a glow. In the arc positive column the average energy of the electrons is from one-half to one-third the energy of electrons in the positive column of a glow and is in the range of 1 to 3 volts. This means that although many of the electrons have an energy sufficient to ionize directly such reaction products as NO, CO, and CN in a high-pressure carbon arc, for example, their energy is not sufficient to ionize molecular or atomic hydrogen, nitrogen, and oxygen except by successive impacts. Ionization by successive impacts was not an important source of ionization in the glow; but since the ion densities encountered in the arc, approximately 10^{20} per cu. m., are from 10^4 to 10^5 times as large as in the glow, ionization by successive impacts assumes considerable impor-

¹ Exceptions occur in the continuous transition from glow to arc and in the case of the cathode of a thermionic arc heated by external means. In the latter case the cathode fall may be considerably less than 10 volts.

² There is little difference in the positive-column characteristics of low-pressure arc and glow discharges (see Secs. 10-4 and 10-6).

tance. Because of the high density of electrons a considerable amount of energy is transferred to neutral atoms in elastic collisions. This has the effect of raising the gas temperature. At pressures of 1 atm. or more the gas temperature attains values of 6000° or more. This is sufficiently high for thermal ionization of the gas to occur. It must be borne in mind, however, that this temperature is really attained because of the motions of a very high density of electrons. In both the glow and the arc, it is the motions of the electrons that ultimately produce the ionization in the positive column; but though in the glow discharge the electrons ionize by direct impact, in the arc they produce ionization indirectly by transferring energy to neutral molecules which in turn produce ionization by collision after acquiring sufficient energy from successive electron impacts. The positive-column gradient varies widely in various gases, but it always assumes a value just sufficient to provide for the regeneration of those ions and electrons which are lost by diffusion into the cooler gases surrounding the arc and by recombination and to supply the heat losses from the positive column by radiation and conduction. In most cases the gradient is quite small, and the case of the copper arc cited above is typical.

Whereas the anode and cathode falls are almost independent of the arc current, the drop in the positive column depends markedly on the magnitude of the arc current. This can be illustrated by means of the results obtained from studies of the carbon arc.

The carbon arc in air has probably been studied more extensively than any other type of arc discharge, but it is hardly the simplest discharge to deal with in a quantitative manner. The pole pieces are constantly boiling or burning away, reacting with the gas to form complex oxides, and creating a great deal of turbulence in the arc region. In addition, convection currents of air are continually changing the arc length. However, since the carbon arc is so familiar to everyone, the results for that material will be quoted. As a rule, the drop increases uniformly with arc length according to the relation

$$E_p = A + Bl \quad (11-1)$$

in which A and B depend on the current. It is natural to assume that the constant A in this equation corresponds to the sum of

the anode and cathode falls and that the product Bl represents the uniform drop in the positive column. This is essentially true. Apparently, all the increase in arc drop with arc length can be accounted for in the positive column.

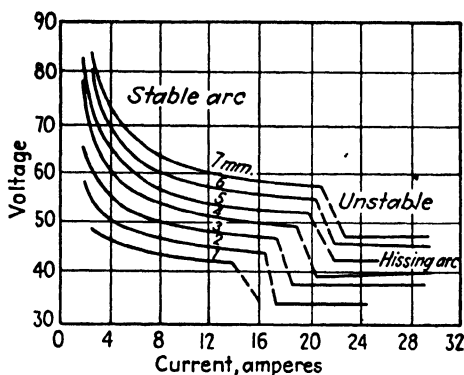


FIG. 11-5.—Volt-ampere characteristics of a carbon arc. (H. Ayrton, *The Electrician*, 35, 419, 1895.)

As for the variation of voltage with current at constant arc length, typical curves are shown in Fig. 11-5. Here the characteristic corresponds to the equation

$$E_p = C + \frac{D}{I} \quad (11-2)$$

on the left of the break. At the break in the curves the type of arc changes from a normal arc to an abnormal, or hissing, arc. Multiplying both sides of Eq. (11-2) by I , we obtain

$$E_p I = CI + D \quad (11-3)$$

which tells us that the power input to the arc is a linear function of the arc current. Combining Eqs. (11-1) and (11-2), we can get a more general expression for the dependence of the arc drop in the carbon arc on both the current and the arc length, as follows:

$$E_p = a + bl + \frac{c + dl}{I} \quad (11-4)$$

This relation is known as Ayrton's formula, having been named for Mrs. H. Ayrton who made and published probably the first

extensive study of arc characteristics.¹ Particular values for the constants a , b , c , and d have no importance, for they vary considerably, depending on the experimental conditions of the test and the kind of carbons used for the electrodes. However, the order of magnitude of the constants is indicated by the values 38.9, 2,000, 16.6, and 10,500 which apply to the constants a , b , c , and d , respectively, under one set of conditions. Obviously, Eq. (11-4) holds only in the limited range of the stable arc indicated by Fig. 11-5; for at very small values of current the arc changes to a glow, and at very large values of current it changes into an abnormal state characterized usually by a distinct hissing sound. This hissing sound seems to be due to fluctuations in gas emission from the anode at high current densities. Apparently, when the current density exceeds a certain value, the vapor emission from the anode becomes so large that the arc is cooled; a reduction is thus caused in the anode fall of potential and consequently in the voltage necessary to maintain the arc, but at the same time this induces an unstable condition which leads to fluctuations in current density and gas emission from the anode.

Using electrodes of any material other than carbon, one might at first expect the same sort of volt-ampere characteristics. Actually, the characteristics are similar, but they do not agree completely with Eqs. (11-2) and (11-3). They do, however, follow the more general relation

$$E_p = C I^n \quad (11-5)$$

where n has various values depending upon the electrodes used and the gas or vapor in which the arc is drawn. In the case of short arcs in air, Nottingham² has found that the exponent n is a linear function of the boiling point of the anode material. Apparently, for the case of a short arc in air at atmospheric pressure where the anode fall of potential is high, the temperature of the anode increases to the boiling point of the material of which it is composed and the vapor thus evolved is sufficient so that the arc actually exists in the vapor of the anode material. The arc characteristic would thus depend strongly on the type

¹ "The Electric Arc," The Electrician Publishing and Printing Co., London, 1902. See also earlier work by the same author.

² *Phys. Rev.*, **28**, 764, 1926.

or kind of electrodes used. This has been verified by determining the volt-ampere characteristic of an arc with cooled electrodes (cooled below the melting point, at least). In those cases the characteristic was determined, even for very short arcs, by the kind of *gas* between the electrodes rather than the kind of electrodes used. It is strange that the one element, carbon, for which the earliest and most extensive determinations of arc characteristics were made, happened to have a boiling point at just such a value that the exponent of I in Eq. (11-5) turned out to be unity.

From the similarity in the glow characteristic and the arc characteristic, it is obvious that all the discussion of stability in glow discharges in Sec. 10-8 will apply equally well to the arc. Just as in the case of a glow discharge, there are two points of equilibrium for the discharge. That is, it is possible to have the arc voltage plus the external-resistance drop equal to the impressed voltage at two values of current. However, at the lower of the two currents the arc is unstable. If at the lower current point of equilibrium the current should drop slightly the arc voltage would increase, the sum of arc drop and resistance drop would exceed the circuit voltage, and the current would decrease further, the process being repeated until finally the discharge changes to a glow or the current drops to zero and the arc goes out. On the other hand, if the current should increase slightly, the arc drop would decrease, the sum of arc drop and resistance drop would be less than the circuit voltage, and the current would continue to increase until a condition is again reached where arc voltage plus resistance voltage is equal to circuit voltage. At the higher current point of equilibrium the arc is stable; for there, after small increases or decreases in current, the arc tends to return to its original condition.

11-3. The Arc Positive Column.—The fact that an unconfined arc in air takes a definite cross section is familiar to everyone. Unlike a glow discharge which appears to fill the entire containing vessel, the arc at high pressure appears to occupy only a very small volume. However, even a glow discharge does not always fill its confining tube. As the pressure is raised in the glow discharge, the glow gradually pulls away from the walls and contracts into a path more or less near the center of the discharge tube. The situation is very much the same in the case of arcs.

At pressures of 1 mm. or less the discharge fills the entire container; but as the pressure is raised, the discharge pulls away from the walls, and at pressures of the order of 1 atm. or more the arc occupies only a restricted path and has a very definite visible boundary. An example of this transition occurs in the high-intensity mercury-arc lamp. In this the discharge is started at reduced pressure; but as the tube heats up, the mercury vapor



FIG. 11-6.—Wet flashover on a six-unit string of suspension insulators. (*Courtesy of Ohio Brass Co.*)

pressure increases, and at its eventual operating condition the arc only partly fills the tube, occupying a path of definite cross section near the center of the tube. In the breakdown of a spark gap or the flashover of an insulator string with alternating voltage (see Fig. 11-6), the arc often seems to occupy many distinct paths in parallel and photographs of such breakdowns exhibit a lacework of paths taken by the arc. In such cases,

however, the reason why we observe so many paths is that during the time that the voltage is reversing and the arc is out the hot gases composing the arc core move slightly. On the next half cycle, then, the arc appears to break down in a different path.

In the low-pressure arc, the positive column differs very little from the positive column of a glow discharge, the only difference being that the current density is usually somewhat greater than in the glow. The ionizing and deionizing agents active in the positive column of a low-pressure arc and the methods used to determine its characteristics are all the same as in the glow. Because of this similarity, we shall discuss here only the characteristics of the positive column of high-pressure arcs (*i.e.*, pressures of the order of 1 to 10 atm.) and leave it to the reader to review the material in the preceding chapter on the positive column of the glow, bearing in mind that it applies to the arc as well.

In the high-pressure arc, certain characteristics of the positive column stand out in all experiments. These must be accounted for by any adequate theory. In the first place the potential gradient in the positive column is always low although not necessarily so low as in the positive column of the glow discharge. Furthermore, the potential gradient decreases with an increase in arc current. The fact that the arc has a definite cross section was mentioned above, but the arc cross section is apparently independent of the length of the arc and depends only on the current flowing. The temperature or average energy of the electrons, positive ions, and neutral molecules of the gas is approximately the same in the arc and corresponds to a temperature of about 6200°K. Actually, the gas temperature also seems to depend on the arc current to some extent varying from about 5950°K. at 1 amp. to 6400°K. at 10 amp. in a nitrogen arc at atmospheric pressure, and it may be even higher at larger arc currents. It is quite apparent from all the experiments on stable high-pressure arcs that there is really only one independent variable in the arc and that is the current. Whenever the current changes, the arc readjusts itself to new conditions and we observe that the arc cross section, potential gradient, and gas temperature acquire new values. This is the important distinction between the high-pressure-arc positive column and other kinds of conductors.

Although the diameter of the positive column of a high-pressure arc can be determined easily from visual or photographic measurements, the voltage gradient, gas temperature, and ion density

cannot be obtained so easily. The problem is much more difficult than the determination of positive-column characteristics in a glow discharge, for probe measurements are practically useless. The gas temperature is so high that, unless the probe is very large or is moved very rapidly through the arc, it is quickly destroyed. But even if the probe is large enough or moves quickly enough through the arc, results obtained with it cannot be relied upon because the very introduction of a cold probe into the arc region upsets the equilibrium and introduces errors into the measurements. Since the anode and cathode falls remain practically constant, the gradient in the positive column can be obtained fairly accurately, however, by measuring the change in the total arc drop with changes in arc length. The positive-column gradient can also be measured with a probe if care is taken in the interpretation of results. Spectroscopic means have been used to determine the gas temperature and the ion density, principally by Ornstein¹ and his coworkers; but other estimates of the gas temperature have been made by determining the gas density through measurements of the absorption coefficient of the gas for alpha particles, beta particles, or soft X rays and through measurements of the velocity of sound in the hot gas.²

The problem of accounting for the positive-column characteristics is not an easy one if a rigorous solution is desired, although a qualitative check with experiment can be obtained relatively easily. Essentially the solution depends upon an energy-balance computation; for all the energy input to the positive column goes to maintain a sufficient ion concentration for the arc current, and every ion pair lost from the positive column represents the loss of a definite amount of energy. In other words, the number of ions produced in the arc must be just equal to the number of ions lost to the wall or to the surrounding cool gas. In the glow discharge, practically the only source of ionization in the positive column is ionization by collision of electrons and neutral gas atoms. Now, in the arc discharge, the current densities involved are much larger than in the glow, and hence

¹ ORNSTEIN, L. S., and H. BRINKMAN, *Physica*, **1**, 797, 1934. TER HORST, D. T. J., H. BRINKMAN, and L. S. ORNSTEIN, *Physica*, **2**, 652, 1935.

² V. ENGEL, A., and M. STEENBECK, *Phys. Rev.*, **37**, 1554, 1931; *Wiss. Veroff. Siemens K.*, **12**, 1, 74, 81, 1933. RAMSAUER, C., *Electrotech. u. Maschinenbau*, **51**, 189, 1933. SUITS, C. G., *J. App. Phys.*, **6**, 190, 196, 315, 1935; **10**, 728, 1939; *Proc. Nat. Acad.*, **21**, 48, (1935); *Gen. Elec. Rev.*, **39**, 194, 1936.

the ion densities must be much larger. The increase in ion density with only a slight increase in positive-column gradient can be explained as due, in part, to thermal ionization occurring in the gas of the arc core because of the high energy input into the positive column. Of course, ionization by collision of electrons with neutral atoms still accounts for much of the ionization.

Because of the strong dependence of thermal ionization on gas temperature (see Sec. 8-2), the ion density present at any time will depend greatly on the gas temperature. On the other hand, the energy input to the arc positive column per unit volume per

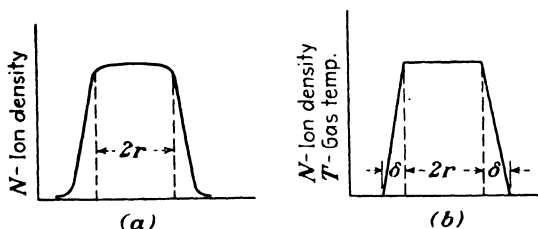


FIG. 11-7.—Approximate ion concentration and gas temperature over the cross section of the arc positive column.

second is equal to the product of the current density and the voltage gradient. But since the current density is given approximately by the electron current density

$$J = N_e q_e K F$$

it is reasonable to suppose that the greatest energy input will occur in those regions in which the electron concentration N_e is the greatest. Qualitatively, then, we can justify a distribution of ion density similar to that indicated in Fig. 11-7a. All this, of course, is based on the assumption that there is no recombination in the arc core and that all the energy losses occur at the boundary of the arc. This is not an unreasonable assumption; we saw before, in Sec. 7-3, that there is very little recombination in a gas at high temperature. Actually, there is probably some recombination going on between electrons and positive ions in the arc core, but the loss in energy by this means is entirely negligible in comparison with the other losses.

Energy is lost from the positive column in no less than six different ways. These are (1) recombination, (2) radiation, (3) radial drift, *i.e.*, drift of ions in the radial electric field if one exists, (4) diffusion, (5) thermal conduction, and (6) convection.

Of these, only the first two can be neglected in most situations. Recombination, as mentioned above, depends very strongly on the gas temperature and begins to become important only in the region outside the arc core. Of course, recombination is eventually responsible for the disappearance of ions in a free arc because diffusion only accounts for the motions of ions from one region of high density to another of low density and, after the ions disappear from the arc core by diffusion, recombination is still necessary to cause their complete annihilation. But as far as influencing arc characteristics is concerned, recombination can be neglected. Although the light emitted from an arc is very intense and although arcs constitute probably the most efficient means of producing light, the fraction of the total energy lost that is lost in radiation is not more than a few per cent. Ornstein and Brinkman¹ give the fraction of the energy input to a carbon arc that is radiated from the positive column as 0.8 per cent at 5 amp., 1.4 per cent at 10 amp., and 3.0 per cent at 14 amp. So, without much inaccuracy, this can be neglected, also.

Of the remaining factors, we can group the loss by radial drift with the loss by diffusion and treat the combination as a single loss due to ambipolar diffusion. Although convection accounts for an appreciable share of the energy loss in the positive column, calculations involving it are difficult and for a first approximation it may be neglected. We are then left with essentially two ways in which energy is lost from the positive column, *i.e.*, diffusion and thermal conduction.

Although the processes involved in thermal conduction and in diffusion are essentially the same—both involving a motion of gas particles out of the region of the arc core—they cannot be treated together. Due to the difference in temperature of the arc core and the surrounding medium, there will be a net transfer of molecules from the arc core to the surrounding volume, each molecule taking with it an energy of about 1 volt. The molecules that move back into the arc core to replace the high-temperature molecules will carry with them only a fraction of a volt of energy, and so there will be a net flow of energy outward from the arc core due to this thermal conduction. In addition to this loss, there is a loss of ionized molecules due to the gradient in the ion concentration near the arc core. Each pair of these ions

¹ *Physica*, 1, 797, 1934.

will carry with them an energy of recombination equal to the ionization potential of the gas—approximately 15 volts in nitrogen. Thus, although the number of particles lost from the arc core by diffusion may be small in comparison with the number lost through thermal conduction owing to the fact that only a fraction of the gas is ionized, the energy lost by diffusion can be almost the same as the energy lost by thermal conduction, for each ion carries ten or more times the energy of a neutral molecule.

Lamar, Stone, and Compton¹ following the original method of Compton² have solved this problem in a rigorous manner, and Stone and Lamar³ have also extended the theory to include convection. The results of their calculation show that they are able to predict the falling volt-ampere characteristic of the arc positive column, as well as the observed temperature variation in the positive column with the arc current; but a discussion of their method is beyond the scope of this volume. All that we shall attempt to show in a qualitative way, following the method of Slepian,⁴ is that the volt-ampere characteristic is a falling one and that the arc cross section is determined by the current flowing in the positive column.

If we let F be the potential gradient in the positive column, then the energy input W per unit length per second will be

$$W = IF \quad (11-6)$$

where I is the total arc current. In order to estimate the losses from the arc, we can assume in this approximation that the ion density is constant over the arc cross section and decreases to zero uniformly in a distance δ outside the arc, as shown in Fig. 11-7*b*. Similarly, we can assume that the arc temperature is also constant over the cross section of the arc core and decreases to zero uniformly in a distance δ outside the arc core, as shown in Fig. 11-7*b*. However, the radius of the arc r is not necessarily constant. We shall assume, though, that it always adjusts itself so that the energy input to the arc, and hence the energy loss, per unit time are a minimum. According to these assump-

¹ *Phys. Rev.*, **55**, 1235, 1939.

² *Phys. Rev.*, **21**, 266, 1923; *J.A.I.E.E.* **46**, 868, 1927.

³ *Phys. Rev.*, **57**, 212, 1940.

⁴ General reference 11-1, pp. 167-168.

tions, the arc positive column can be treated as a uniform, hot conductor whose diameter $2r$ depends upon the current in the conductor. This picture is naturally very approximate; but the conditions assumed here do not differ greatly from actual conditions, and we should be able to predict the trend of the arc characteristics, at least.

Let us say, then, that W_c is the loss in energy from the positive column by thermal conduction per unit length per second and W_d is the loss in energy per unit length per second by diffusion of electrons and positive ions from the arc core to a region of low density outside the arc. Since the net heat flow per second through unit area at any point is proportional to the temperature gradient at that point, the net energy lost from the arc positive column per second is proportional to the circumferential area of the arc and the temperature gradient at the arc boundary. Thus, W_c is given approximately by

$$W_c = A_c 2\pi r \frac{T}{\delta} \quad (11-7)$$

where A_c is the constant of proportionality. Similarly, since the net number of ions diffusing per second through unit area at any point is proportional to the gradient in the ion concentration at that point, we can write

$$W_d = A_d 2\pi r \frac{N}{\delta} \quad (11-8)$$

where A_d is the other constant of proportionality and N is the ion concentration in the arc core. Actually, A_c is proportional to the average velocity, mean free path, and specific heat of the gas molecules; and A_d is proportional to the average velocity and mean free path of the ions. But if we assume for this calculation that the temperature of the gas is independent of the current flowing in the arc, we can treat A_c and A_d as essentially constant. Furthermore, under this condition, W_c will depend only on r . Equations (11-7) and (11-8) can then be simplified to give Eqs. (11-9) and (11-10) if we lump all the constants of each equation into a single new constant. Thus,¹

$$W_c = A'_c r \quad (11-9)$$

$$W_d = A'_d r N \quad (11-10)$$

¹ In this, we have also assumed that δ remains constant as the current is varied.

In Eq. (11-10), the ion concentration N still appears. This cannot be measured directly; but we do know that, since the electrons always have a much higher mobility than the positive ions, the total drift current flowing in the arc is practically all electron current. We can then say that, to a close approximation, the arc current I is given by

$$I = \pi r^2 N_e q_e K_e F \quad (11-11)$$

where K_e is the mobility of the electrons, $K_e F$ is the average drift velocity of the electrons, and $N_e q_e K_e F$ is the average drift current density. Since the concentration of electrons is approximately the same as the concentration of positive ions, Eq. (11-11) can be solved for N_e and the value substituted in Eq. (11-10) to give

$$W_d = \frac{A'_d r I}{\pi r^2 q_e K_e F}$$

or

$$W_d = \frac{A''_d I}{r F} \quad (11-12)$$

since K_e will be essentially constant at constant temperature and can be combined with the other constants to give a new proportionality constant A''_d . We see now that, although the loss by thermal conduction is independent of the positive-column current and potential gradient and depends only on the radius of the positive-column cross section, the loss by diffusion is directly proportional to the arc current and inversely proportional to the positive-column gradient and the radius of the arc core.

The energy-balance equation thus becomes

$$W = W_c + W_d$$

or

$$IF = A'_c r + \frac{A''_d I}{r F} \quad (11-13)$$

In Fig. 11-8, Eq. (11-13) has been analyzed graphically for the condition when F and I are constant. Because the ion density decreases as the square of the arc radius when the total current is maintained constant, the total energy lost by diffusion actually

decreases as r increases. This is in agreement with the conclusions reached in Sec. 7-2. There we saw that, when the only deionizing agent was diffusion and a uniform volume source of ionization was maintained, the relative loss in ionization for a given total ionization decreased as the radius of the discharge tube was increased. Here, the dependence on r is not quite the same as in that problem, but we reach the same conclusion. As shown by the graph, the loss by thermal conduction increases directly with the arc radius. The total loss, consequently, increases as r is decreased for small values of r owing to the increased loss by diffusion, and the total loss increases as r is increased for large values of r owing to the increased loss by thermal conduction. Thus, there is a certain value of r for which the loss in energy is a minimum. This determines the actual arc radius, for conditions will adjust themselves so that the loss in energy from the arc is a minimum.

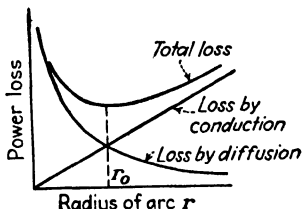


FIG. 11-8.—Power loss in the arc.

If we solve Eq. (11-13) for F , we obtain

$$F = \frac{A'_c r}{2I} + \sqrt{\left(\frac{A'_c r}{2I}\right)^2 + \frac{A''_d}{r}} \quad (11-14)$$

and if we multiply this equation by I , we get

$$FI = \frac{A'_c r}{2} + \sqrt{\left(\frac{A'_c r}{2}\right)^2 + \frac{A''_d I^2}{r}} \quad (11-15)$$

Equation (11-15) also shows that the power input, with a given current, will be a minimum for a particular value of r and the arc should have a definite cross section. Equation (11-14) shows that the gradient in the positive column, and consequently the arc voltage when the length of the arc is constant, will decrease with increasing current. In these two important respects, then, the approximate theory agrees with experiment. Actually, the arc does not have an absolutely sharp boundary, as this calculation might indicate, the ion density and the gas temperature falling off gradually in the gas surrounding the arc core as shown in Fig. 11-7a. But since thermal ionization depends so strongly on the gas temperature, and most of the current is

carried within a limited volume, for all practical purposes, one can speak of the arc as having a definite boundary or cross-sectional area.

If an arc is confined in a small tube of refractory material, it is evident from the fact that the arc assumes a definite size for every value of arc current that the size of the tube will make no difference in the voltage drop along the arc until the current in the arc is increased to the point where the natural, or free, arc diameter tends to exceed the diameter of the tube. Since the arc diameter cannot exceed the diameter of the hole it must remain constant with any further increase in current. This means, then, that as the arc current increases, the total loss in energy from the arc corresponds to points farther and farther to the left of the minimum of the total loss curve in Fig. 11-8, for the free arc would have a larger and larger diameter. Consequently, the loss of ionization by diffusion increases rapidly with increases in current, and the voltage gradient necessary to maintain the discharge must increase rapidly with current. This behavior has been verified by experiment. When an arc is drawn in a small hole in soapstone, for example, it is found that as the current is increased the voltage gradient at first decreases but after the arc completely fills the hole the voltage gradient increases rapidly. :

One important application of this characteristic of high-pressure arcs is in the porous-block lightning arrester.¹ As mentioned before in Sec. 10-9, an ideal lightning arrester is one which permits a large current to flow with no decrease in voltage, *i.e.*, one for which the volt-ampere curve is a horizontal line. If a refractory block containing many small pores is used as a lightning arrester, the breakdown voltage of the block is unaffected by the size of the holes; but once a discharge is started in one hole, the voltage decreases only until the arc fills the hole, and then the drop across the block increases with any further increase in current. With a proper porous material, this arc drop can very soon be made to exceed the breakdown voltage; then, as the current increases, all that happens is that more and more pores break down and carry current, but the arc voltage does not change. As the current falls to zero, the arc

¹ SLEPIAN, J., R. TANBERG, and C. E. KRAUSE, *Trans. A.I.E.E.*, **49**, 257, 1930.

is quickly extinguished; or if the voltage falls below the breakdown voltage, the discharge is extinguished.

It may be mentioned here that, although lightning arresters are quite generally used to protect equipment and transmission lines from the effects of surges due to lightning disturbances, they are not completely successful; *i.e.*, commercial arresters are not built to protect equipment against the most severe surge conditions such as occur in direct lightning strokes. However, they can be built to protect against most of the surges appearing on a line.

11-4. The Cathode of the Arc.—A study of the manner in which the transition from glow to arc occurs reveals that apparently we cannot lay down a single all-inclusive theory of the arc cathode. We shall have to define at least two kinds of self-maintaining arc that depend on the material of the cathode. In the case of high-melting-point materials, such as tungsten or carbon, the cathode in all probability attains a temperature high enough to emit electrons thermionically; if the whole cathode is not hot enough for thermionic emission, at least some part is. On the other hand, self-maintaining arcs are quite frequently observed when the cathode is a material such as mercury which boils at a temperature far below that necessary for thermionic emission. For these arcs, another mechanism has to be devised. More will be said about this a little later. With either kind of arc the local spot from which the arc appears to spring is called the *cathode spot*. The size of this spot is difficult to ascertain because of its extreme brightness, but experiment seems to indicate that it varies with the arc current, the relationship being approximately linear.

The temperature of the hottest part of the cathode of a tungsten arc has been found to be approximately 3300°K. when the cathode current density was about 0.5 amp./sq. mm. Since in the neighborhood of 3300°K. a very slight change in temperature results in a large change in emission and since between 0.1 and 0.2 of the cathode current is positive-ion current, the correspondence with thermionic-emission data as given in Table 11-1 is quite close. In the case of carbon arcs the cathode-spot current density varies from 2.0 to 5.0 amp./sq. mm.; but in these cases the cathode temperature is about 3600°K., and so this also agrees with thermionic-emission data.

TABLE 11-1.—ELECTRON EMISSION FROM TUNGSTEN

Temperature, deg. K.	Electron Emission, amp./sq. mm.
2000.....	0.00001
2400.....	0.00116
2800.....	0.0354
3200.....	0.478
3300.....	0.844
3400.....	1.42
3600.....	3.73
3655.....	4.79

That ordinary thermionic emission is sufficient to explain the maintainance of a low-voltage hot-cathode arc discharge is

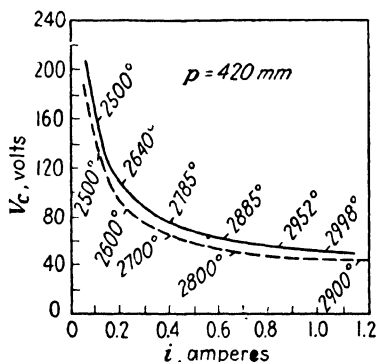


FIG. 11-9.—Glow-to-arc transition with tungsten electrodes in nitrogen. (General reference 11-5, p. 142.)

illustrated by the fact that it is possible to compute almost exactly the voltage-current characteristic of the transition from glow to arc when a tungsten cathode is used. Figure 11-9 shows the results of the calculation. The solid curve represents the experimental data¹ and the dotted curve the theoretical computation.² The method employed in the calculation of the dotted curve is essentially the following: The

total current i at the cathode is made up of electron current i_e and positive-ion current i_p , thus:

$$i = i_e + i_p \quad (11-16)$$

But the electron current also can be considered to be made up of two parts, namely, a part due to thermionic emission, i_{th} , and a part due to secondary-electron emission by the incoming positive ions, γi_p ,

$$i_e = i_{th} + \gamma i_p \quad (11-17)$$

These two equations can now be solved to give the electron and positive-ion currents in terms of the total current and the thermionic-emission current, as follows:

¹ BÄCHTIGER, P., *Helvetica Phys. Acta*, 3, 335, 1930.

² General reference 11-5, p. 142.

$$i_e = \frac{\gamma i + i_{th}}{1 + \gamma} \quad (11-18)$$

$$i_p = \frac{i - i_{th}}{1 + \gamma} \quad (11-19)$$

The number of positive ions per meter produced by each electron leaving the cathode is given by Eq. (8-4) of Sec. 8-3; hence, the total ionization produced in 1 sec. by all the electrons leaving the cathode is $P(V_c - E_i)i_e$, where V_c is the cathode fall of potential, E_i is the ionization potential of the gas, and P is a new constant of proportionality. If we consider that a fraction k of the positive ions, so produced, reach the cathode, then i_p is given by

$$i_p = kP(V_c - E_i)i_e \quad (11-20)$$

or the ratio of i_p to i_e is

$$\frac{i_p}{i_e} = kP(V_c - E_i) \quad (11-21)$$

Substituting the values above for i_p and i_e , we obtain

$$kP(V_c - E_i) = \frac{i - i_{th}}{\gamma i + i_{th}} \quad (11-22)$$

Now, at the cathode, energy is lost by radiation and by the conduction of heat through the cathode support and through the gas. The loss of heat by radiation is proportional to the fourth power of the cathode temperature, and the loss by conduction is proportional to the first power of T ; so we can write

$$\text{Energy loss} = BT^4 + CT \quad (11-23)$$

In equilibrium, this must be equal to the energy input to the cathode, which can be considered to have three principal components. (1) A heating component due to positive ions falling through the cathode-potential drop and giving up their energy to the cathode. If f represents the fraction of the positive ions giving up energy of this kind to the cathode,¹ the total energy gained will be fi_pV_c . (2) Another heating component due to

¹ $(1 - f)$ will represent the fraction of the positive ions that retain the energy gained in the cathode fall and after neutralization at the cathode pass away as molecules of high kinetic energy. f is called the *accommodation coefficient*.

the neutralization of the positive ions at the cathode. This is equal to $i_p E_i$. (3) A cooling component due to the emission of electrons from the cathode, to supply both i_e and the electrons for neutralizing the incoming positive ions. This is given by $i\phi$ where ϕ is the work function of the cathode surface. Thus the energy input to the cathode is equal to $(fi_p V_c + i_p E_i - i\phi)$. Equating this to the energy loss, given by Eq. (11-23), we have

$$i_p(fV_c + E_i) - i\phi = BT^4 + CT \quad (11-24)$$

or, substituting the value of i_p from Eq. (11-19),

$$\frac{i - i_{th}}{1 + \gamma} (fV_c + E_i) - i\phi = BT^4 + CT \quad (11-25)$$

It is possible now to combine Eqs. (11-25) and (11-22), and, remembering that $i_{th} = AT^2 e^{-\frac{b}{T}}$, to eliminate both i_{th} and T and obtain an equation relating V_c to the current i . This is the relationship that has been plotted in Fig. 11-9 as the dotted curve.

As a comparison of Eqs. (11-18) and (11-19) shows, the positive-ion current at the cathode of a tungsten arc may be small in comparison with the electron current, especially when the cathode temperature is high. Experimentally, i_p is found to be about 0.1 of the total current; so in this respect, also, the theory seems to be correct.

When the cathode of an arc is heated by external means as it is in a phanotron, the phenomena taking place at the cathode are essentially the same as in the hot-cathode arcs described above. The only difference is that the cathode fall of potential is usually less in the non-self-maintaining arc than it is in the self-maintaining arc, for the cathode is maintained hot enough for thermionic emission by external means. The thermionic emission, however, is of the same order of magnitude relative to the total current as in the self-maintaining thermionic arc, and the cathode fall is approximately equal to the ionization potential of the gas.

Thermionic emission does not explain all self-maintaining arcs by any means. It is quite evidently adequate to explain the existence of arcs between tungsten or carbon electrodes, for these can be easily heated to temperatures high enough for thermionic emission to supply practically all the current. But

consider the other metals in Table 11-2. Arcs with any of these metals are familiar. Yet most of them boil at temperatures considerably below that required for an appreciable amount of thermionic emission (about $3000^{\circ}\text{C}.$). There are other differences, too, that distinguish an arc employing a tungsten or carbon cathode from one with a so-called "cold cathode" such as mercury, aluminum, or zinc. Whereas the cathode spot

TABLE 11-2

Metal	Melting point, deg. C.	Boiling point, deg. C.
Hg	— 38.9	356
Cd	320.9	766
Na	97.5	880
Zn	419.5	907
Al [•]	659.7	1800
Cu	1083.0	2300
Fe	1535.0	3000
W	3370.0	5900
C	3500.0	4200

always remains stationary on the surface of a carbon or tungsten cathode, with a cold-cathode arc it frequently races about over the surface of the cathode at speeds estimated at 0.5 to 1.5 m./sec. and even as high as 10 m./sec. With these surfaces, too, although the size of the cathode spot is roughly proportional to the cathode current, as with thermionic arcs, the current densities are very much larger than with a thermionic arc. Estimates of the current density vary from 17 to 180 amp./sq. mm., compared with 0.5 to 5 amp./sq. mm. in the cathode spot of the thermionic arc. Furthermore, the potential drop at the cathode is much less, often just exceeding the first critical potential of the gas and much lower than with a thermionic arc. Finally, unlike thermionic arcs, which can be reignited easily after an interruption of almost 1 sec., the cold-cathode arc cannot be reignited in many cases when the interruption is as short as 10^{-9} sec. Because of these differences, we are forced to assume that there are at least two distinct types of self-maintaining arc, those with high-melting-point cathodes, which we shall call *thermionic* or *hot-cathode* arcs, and those with cathodes that boil at temperatures

insufficient for thermionic emission, which we shall call *cold-cathode* arcs. The distinction between these two types will be further emphasized when we consider the characteristics of a.c. arcs in Sec. 11-7.

The suggestion that these cold-cathode arcs are maintained by high-field emission is due to Langmuir.¹ He postulated that a space-charge sheath of positive ions forms so close to the cathode surface that the electric intensity at the cathode surface, due simply to the electric field, is sufficient for large numbers of electrons to be emitted. If this is to be true, the electric intensity at the cathode surface would have to be of the order of 10^{10} volts/m. Actually, with the enormously large current densities observed in the cathode spot of a cold-cathode arc, the positive-ion space-charge sheath at the cathode² is so dense that electric intensities of 1.7×10^9 volts/m. have been calculated.³ This is a large enough value so that, allowing for irregularities on the cathode surface that can easily multiply the electric intensity by a factor of 10, electron current densities of the order of 10 amp./sq. mm. can occur.

Further evidence for a cold-cathode effect is found in the fact that the temperature of the surface under the cathode spot in a mercury arc has been estimated at between 250 and 350°C. from measurements of the evaporation taking place. Temperatures for the cathode-spot area of 2000°C. obtained from spectroscopic measurements have been shown to be in error, for in those measurements the radiation was actually coming from the gas and not the mercury surface. Elaborate explanations have been presented by many authors in an attempt to show that the mechanism at the cathode spot in a cold-cathode arc is the same as for a thermionic arc. These arguments have all broken down, however, when all the characteristics of cold-cathode arcs have been considered. J. J. Thomson,⁴ for example, shows that time lags of breakdown as short as 5 microsec. can be explained on the basis of thermionic emission in the ordinary sense when the stream of incoming positive ions prevents the evaporation of

¹ *Gen. Elec. Rev.*, **23**, 735, 1923.

² A sheath similar in all respects to the sheath at the cathode of a glow discharge.

³ General reference 11-5, p. 152.

⁴ General reference 11-2, p. 595.

the cathode material. However, experimentally, time lags of breakdown of 10^{-7} or 10^{-8} sec. have often been obtained, and these are much too short to be explained on any thermionic theory. Furthermore, interruption of the arc for as little as 10^{-9} sec. is sufficient for the cathode spot to disappear and for reignition of the arc to be prevented. It is difficult to see, on the basis of a thermionic theory, how enough cooling could occur in 10^{-9} sec. to prevent the reignition of the arc. In addition to these experiments, tests with Deion circuit breakers have shown that it is possible to split a long arc into many short arcs between copper plates and by means of a magnetic field to move the arc rapidly across the metal plates. Examination of these plates after the cathode spot has traversed them has revealed no evidence of burning or even melting of the copper surface. Certainly, one or the other would have occurred if thermionic emission had been present. The truth probably lies in the fact that no single process can be taken for a complete explanation of the cold-cathode arc. Since the current densities and densities of ionization involved are so very much larger than are ordinarily encountered and since the volume of cathode associated with the emission process is so very much smaller than is usually studied, we cannot say definitely that the usual theories continue to apply. The actual mechanism of electron emission at the cathode spot may be a combination of several familiar mechanisms, or it may be an entirely new one; but until the situation is clarified by further experiment, we can only say that the cold-cathode arc is distinctly different from the thermionic arc and at present *seems* to be made possible by high-field emission.

One other point mentioned above as characteristic of cold-cathode arcs is deserving of attention. In the case of thermionic arcs the cathode fall is always found to be at least equal to the ionization potential of the gas. With cold-cathode arcs, cathode falls only slightly greater than the first critical potential have been observed. This is assumed to be made possible by the fact that with the large current density and large ion concentration present, ionization by successive impacts is taking place and the electrons need have only sufficient energy to produce excitation.

It often happens that the presence or absence of impurities in the cathode metal will determine whether the arc is to be classed

as a thermionic or a cold-cathode arc. It is easy to see that, if a metal such as copper contains alkali impurities, the low work function of the impurity may make a thermionic arc possible even though such an arc is not possible with pure copper. Experiments have shown that the presence of oxygen or an oxidized cathode also facilitates the formation of a thermionic arc, especially when the arc is not moved by external means. Here the effect of the oxide is probably to raise the boiling point to such a degree that thermionic emission can occur. On the other hand, it is possible to have a cold-cathode arc with both tungsten and carbon cathodes under the proper conditions. Evidence for

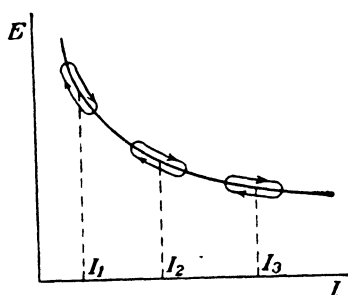


FIG. 11-10.

a cathode of this form is found when breakdown occurs between carbon electrodes in high vacuum; but after the arc forms, it rapidly changes over to a thermionic arc.

11-5. Dynamic Arc Characteristics.—From what has been said about the characteristics of the positive column of an arc, we realize that every change in current in the arc is accompanied

by a complete change in the character of the arc. The gas temperature, ion density, ion and electron energy, cross-sectional area, and gradient in the positive column all depend upon the arc current; and the various points on an arc characteristic such as are shown in Fig. 11-5 actually correspond to different conditions of the arc. If conditions in the arc did not vary with the current, the arc would act like any other conductor and its volt-ampere characteristic would be a straight line through the origin. The changes that do occur, however, cannot occur instantaneously; consequently, we cannot expect the dynamic characteristic, obtained when the current is varied rapidly, to agree with the static characteristics that we have been discussing thus far. As experiments show, the dynamic characteristics of an arc only *approximate* the static characteristics at any time. This seems entirely reasonable when we realize that a change in ion concentration, gas temperature, or arc cross section takes time and involves the input or loss of a certain amount of energy from the positive column.

The dynamic volt-ampere characteristic of an arc can be investigated experimentally by impressing a small varying current on the steady direct current flowing in any test. Figure 11-10 illustrates the results of such a test. When the current is changed very slowly, the static volt-ampere curve is followed exactly; but when the frequency of the pulsations in current are increased, deviations from the static characteristic begin to be observed. The voltage change appears to lag behind the change in current, and a point originally on the static characteristic moves on an ellipse. However, as long as the frequency of the pulsations is small, the ellipse conforms closely to the shape of the static characteristic in the neighborhood of the average value of current. As the frequency of the current pulsations is increased, both the size and the orientation of the ellipse change. The change in the dynamic characteristic is indicated in Fig. 11-11, where the curves aa' , bb' , and cc' correspond to the dynamic characteristic when successively higher frequencies of current pulsation are applied to the arc. It is quite evident that

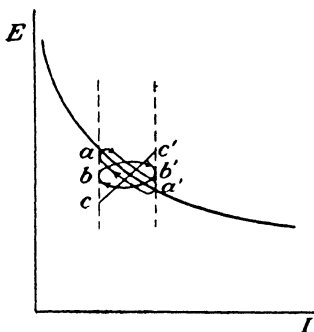


FIG. 11-11.

with the increase in frequency the ellipse rotates counterclockwise and at very high frequencies flattens down to a line of definitely positive slope. We see, then, that, if the arc current can be changed rapidly enough, conditions in the arc core will remain essentially constant and the arc will act like a metallic conductor. *It is only because conditions in a static arc change with every change in current that we get the familiar falling volt-ampere characteristic.*

The hysteresis effect in the arc is always accompanied by a change in the energy balance in the arc; for during the time that the current is increasing, additional ionization is produced which is lost by radial diffusion when the current decreases. The cycle of processes is apparently somewhat as follows: When a sudden increase in current occurs, the voltage increases at first because there is insufficient ionization present to carry the current at the same voltage. Then, with the corresponding increase in energy input in the positive column, its temperature and ion density begin to rise and the cross-sectional area changes to produce

as low an energy input as possible. When these changes occur, the arc voltage drops. As the current is decreased, just the opposite occurs; a larger than necessary ion density, temperature, and cross-sectional area are available, and the arc voltage drops or stays below the static characteristic until the excess ionization is lost by diffusion and the arc temperature has been decreased by thermal conduction. In this way the ellipse is traced out. When the frequency of the current pulsations is raised, the inability of the arc to follow the fluctuations becomes more pronounced and eventually, when the frequency reaches about

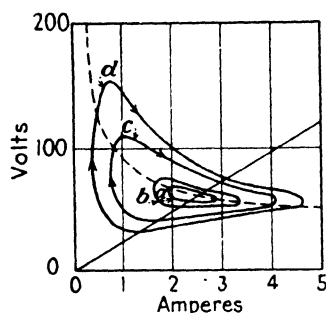


FIG. 11-12.—Dynamic characteristic of a pulsating current arc. (*H. T. Simon, Physik. Z.*, 7, 433, 1906.)

100,000 cycles/sec., the ellipse flattens down to a straight line and the arc acts like a pure resistance. There are differences, however, between the arc and a pure resistance. The slope of the straight-line arc characteristic changes as the average value of current is changed. If the current fluctuates about a large value of current, the slope of the line is less than if the fluctuations occur about a small value of current. Furthermore, the slope of the line multiplied by the current does not

give the corresponding voltage as it would in the case of a metallic resistance. As the amplitude of current pulsations is increased, the deviations from the static characteristic become more pronounced even at relatively low frequencies. Figure 11-12 illustrates this condition. In the figure the static volt-ampere curve is shown dotted and the dynamic characteristics with successively larger currents are shown in the solid lines. All these curves correspond to a frequency of current pulsations of about 450 cycles/sec.

11-6. The Arc Oscillator.—One of the examples of the application of the falling volt-ampere characteristic of an arc is in the design of arc oscillators. Except for the fact that in an arc the currents involved are much larger and consequently the power available to an oscillating circuit is much larger, there is little difference between an arc oscillator and a glow-tube oscillator such as is discussed in Sec. 10-9. The fact that the

dynamic volt-ampere curve of an arc as discussed in the preceding section is not the same as the static characteristic limits the applicability of this type of oscillator, however, to relatively low frequencies where the characteristic still has a negative slope. This means that ordinarily an arc oscillator cannot be used for frequencies much greater than about 50,000 cycles/sec.

As in the case of glow oscillators, there are two types of arc oscillator. Figure 11-13 shows the type of circuit that would be used with the intermittent-arc oscillator. A resistance R is connected in series with the spark gap G , and the resistance is made large enough so that a stable arc cannot be maintained with the voltage available. If E is larger than the sparking potential of the gap, the potential across G will rise slowly owing to the condenser in parallel until the voltage reaches the sparking potential.

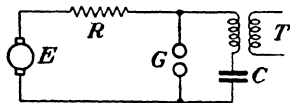


Fig. 11-13.—Intermittent arc oscillator.

Then a breakdown to an arc occurs, and the condenser discharges through the gap. Because the current from E is limited to a value insufficient for a stable arc, the arc goes out after the condenser discharges, the voltage across G and C again starts to rise, and the cycle is repeated. By this means, then, an oscillating voltage can be obtained from the transformer T , which is connected in series with the condenser. Because of the resistance in series with the power source, this circuit is, of course, limited to relatively low-power applications.

That an arc cannot be maintained if R exceeds a certain value is evident from the analysis of the conditions for stability in Sec. 10-8. This can also be verified analytically if we assume that the arc voltage is given by a relation of the form of Eq. (11-5)

$$E_p = C + \frac{D}{I^n} \quad (11-5)$$

Here C and D are constants for a particular arc. Combining this with the equation for stability [Eq. (10-15)], we get

$$E_p = E_b - IR = C + \frac{D}{I^n} \quad (11-26)$$

which can be solved for R to give

$$R = \frac{E_b - C}{I} - \frac{D}{I^{n+1}} \quad (11-27)$$

Now, if there is a maximum value of R that will give a stable arc current, it can be determined from the condition that, at the maximum, $dR/dI = 0$.

Differentiating Eq. (11-27) with respect to I and setting the result equal to zero, we find that

$$I = \left[\frac{(n+1)D}{(E_b - C)} \right]^{\frac{1}{n}} \quad (11-28)$$

This value of I can now be substituted in Eq. (11-27) to determine the maximum value for R . The result is

$$R_{\max} = \frac{n}{D^{\frac{1}{n}}} \left[\frac{E_b - C}{n+1} \right]^{\frac{n+1}{n}} \quad (11-29)$$

This tells us that as long as R does not exceed this value a stable arc can be maintained. In the case of the carbon arc where $n = 1$, it is impossible, then, to maintain the discharge when R is greater than $\frac{(E_b - C)^2}{4D}$. Equation (11-29) also tells us that when R is fixed a stable carbon arc cannot be maintained if E_b is less than $C + 2\sqrt{DR}$.

The other type of oscillating circuit offers more theoretical interest, for it makes use of the falling volt-ampere characteristic

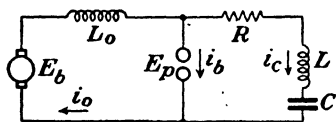


FIG. 11-14.—Continuous arc oscillator.

of the arc to produce sustained pulsations in the arc current. The circuit for this is represented in Fig. 11-14, and the explanation of its operation is similar to that for the glow-tube oscillator described on

page 350. It is easy to show mathematically that undamped oscillations can be obtained from this circuit under the proper conditions. Let us assume that the inductance in series with the supply generator is large enough to prevent any rapid fluctuations in i_0 . Then applying Kirchhoff's laws to the circuit involving R , L , C , and the arc, we have

$$i_0 = i_b + i_c \quad (11-30)$$

and

$$E_p = L \frac{di_c}{dt} + Ri_c + \frac{q}{C} \quad (11-31)$$

where q is the instantaneous charge on the condenser. Differentiating Eq. (11-31) with respect to t , we get, since $dq/dt = i_c$,

$$\frac{dE_p}{dt} = L \frac{d^2i_c}{dt^2} + R \frac{di_c}{dt} + \frac{i_c}{C} \quad (11-32)$$

But since i_0 is to be held constant, Eq. (11-30) says that

$$\frac{di_b}{dt} = - \frac{di_c}{dt} \quad (11-33)$$

Thus,

$$\frac{dE_p}{dt} = \frac{dE_p}{di_b} \frac{di_b}{dt} = - \frac{dE_p}{di_b} \frac{di_c}{dt} \quad (11-34)$$

Substituting this value of dE_p/dt into Eq. (11-32), we get

$$- \frac{dE_p}{di_b} \frac{di_c}{dt} = L \frac{d^2 i_c}{dt^2} + R \frac{di_c}{dt} + \frac{i_c}{C} \quad (11-35)$$

Now here dE_p/di_b is the slope of the volt-ampere arc characteristic at the current i_b ; and since the arc has a falling volt-ampere characteristic, dE_p/di_b is always negative. Thus, if we can adjust R so that

$$R = - \frac{dE_p}{di_b} \quad (11-36)$$

then Eq. (11-35) reduces to

$$0 = L \frac{d^2 i_c}{dt^2} + \frac{i_c}{C} \quad (11-37)$$

The general solution of this equation is

$$i_c = A \sin (\omega t + \theta) \quad (11-38)$$

where A and θ are integration constants that must be evaluated from the experimental conditions and ω has the value $(1/\sqrt{LC})$.

From this we see that the effect of the negative dynamic resistance of the arc is effectively to cancel the damping resistance in the circuit of L and C and allow undamped oscillations to be set up in the arc and its parallel circuit. The frequency of these oscillations is approximately $\frac{1}{2\pi\sqrt{LC}}$. However, the frequency cannot be obtained exactly from the circuit constants, for the discussion in the preceding section has shown that the arc does not follow the static volt-ampere characteristic exactly but instead traces out a loop whose orientation depends upon the frequency of the oscillations. Effectively, the arc behaves as if it possessed inductance, and the actual oscillation frequency is always less than that calculated from circuit elements. Further-

more, since the dynamic arc characteristic changes with the frequency of the pulsations in arc current, it is not possible to obtain a frequency with an arc in air of much over 5,000 or 10,000 cycles/sec. in this way; for with high frequencies the slope

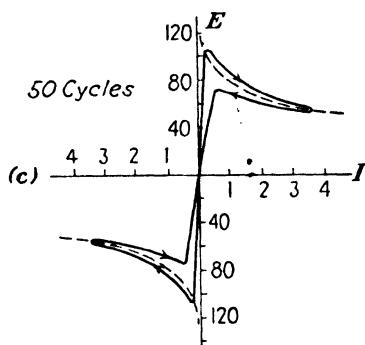
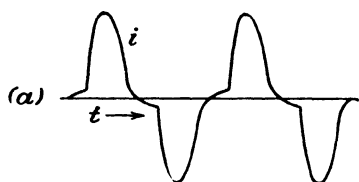


FIG. 11-15.—A.c. carbon-arc characteristics, low current. (*H. T. Simon, Physik. Z.*, 6, 297, 1905.)

of the dynamic characteristic changes from negative to positive, and the circuit resistance cannot be neutralized. In the early days of radio transmission, modifications of this method were extensively used for the generation of radio frequencies. In the Poulsen arc generator, a hydrogen arc is used because thermodynamic equilibrium is attained quicker in hydrogen than in any other gas and consequently the static volt-ampere curve is followed more closely at high frequencies than with any other gas. In addition to this a strong transverse magnetic field was used to produce instability in the arc and to assist in deionization. Even though frequencies as high as 100,000 cycles/sec. have been obtained with the Poulsen arc, it requires such a large magnet and arc chamber for high-power work that it has been superseded by the more convenient and compact vacuum-tube oscillators. Many of the Poulsen arc magnets, which were as large as 50 or 100 tons, have recently been converted into cyclotrons for the generation of high-energy ion beams necessary in the study of nuclear reactions. The Poulsen arc oscillator really operated as a combination of the intermittent- and continuous-arc oscillators, for the fluctuations in arc current were usually so large that the arc current went to zero and the arc was reignited every cycle.

Because of the energy necessary to maintain the arc, the efficiency of the Poulsen arc was inherently low. But efficiencies of 40 to 50 per cent have been obtained under special conditions. This can be compared with efficiencies of about 75 per cent that have been obtained with vacuum-tube oscillators.

11-7. Alternating-current Arc Characteristics.—Deviations from the static volt-ampere characteristic of an arc are most pronounced when alternating current is used. In the case of a 3-mm. carbon arc the dynamic a.c. arc characteristic takes a shape similar to that shown in Fig. 11-15 when the peak current is small and one similar to that in Fig. 11-16 when the peak current is somewhat larger. In both cases the static characteristic is shown as a dotted curve. The dynamic volt-ampere curve follows the static characteristic quite closely when the current is large but deviates considerably when the current approaches zero. The explanation of the curves is quite similar to that applying to small pulsations in current.

As the current is increased, there is a lag in the arc temperature and ionization behind the arc current and the arc voltage must be higher than that required for a constant-current arc in order that the increased ionization can be produced. As the current decreases, the excess ionization does not disappear as rapidly as the current changes and consequently the arc can exist at a lower voltage than that given by the static characteristic. As the current decreases, the arc is not, in the strict sense of the word, self-maintaining, for it depends for its maintenance on the ionization generated at some previous time. As long as the dynamic

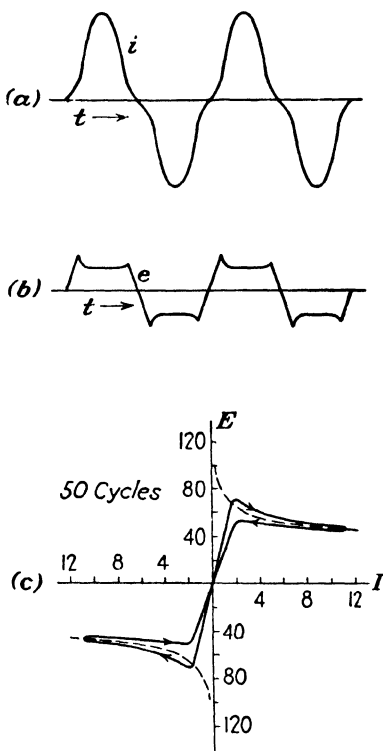


FIG. 11-16.—A.c. carbon-arc characteristics, high current.

characteristic lies above the static characteristic, the rate of ion generation is greater than the rate of ion loss; but when the dynamic curve lies below the static characteristic, the rate of deionization is larger than the rate of ion production. Because of this lag in the change in ionization, the ion density between the electrodes still has an appreciable value when the current and voltage pass through zero. If this were not so, the static characteristic would be followed exactly and the arc voltage would rise to the sparking potential at zero current. In the neighborhood of zero current the arc behaves very much like an ordinary conductor whose resistance is increasing with time. It is only because of this fact that it is possible to reignite the arc on the next half cycle with only a moderate increase in voltage over that necessary to maintain the arc. Reignition occurs easily in a carbon arc anyway, because both electrodes are hot enough to emit large quantities of thermionic electrons and the conductivity of the arc space during the time that the current and voltage are reversing is partly due to these electrons and partly due to the high density of ionization that was necessary to pass the peak current and that has not had time to disappear completely.

For reignition to occur in the carbon arc, the arc voltage must rise high enough so that with the available ionization the dynamic characteristic can cross the static characteristic. Only then is it possible to produce enough ionization in the arc to maintain the discharge. It is easy to see, then, why any treatment that will reduce the arc temperature or ion density in the neighborhood of zero current will increase the voltage necessary for reignition. Thus reducing the a.c. frequency, increasing the external resistance, and increasing the arc length all have the effect of increasing the *reignition voltage*, i.e., the maximum voltage between electrodes necessary to reestablish the arc.

In Figs. 11-15 and 11-16 are also shown curves for the dependence of current and voltage on time. The effect of a high reignition voltage is quite evident. Until reignition occurs, the current remains small; but, because of thermionic emission from the electrodes and the high ion density, it is not zero. After reignition the arc voltage falls rapidly and the current jumps to a high value. As the current decreases, the sine wave is almost exactly followed when the current is large; but with small peak currents the current falls to a small value, and the arc

voltage rises slightly, a short time before reversal occurs. These curves correspond to operation in a 110-volt circuit. Of course, if the circuit voltage is higher, the voltage will rise more rapidly and the reignition voltage will be reached much earlier in the half cycle. Also, conduction will last for a longer portion of the half cycle, for the point at which the circuit voltage falls below the arc characteristic will occur later in the cycle.

When nonthermionic electrodes such as copper are used for an a.c. arc, reignition does not occur so easily. The anode current density of these arcs is never so high as the current density in the cathode spot. Thus, when the potential reverses, the ion density

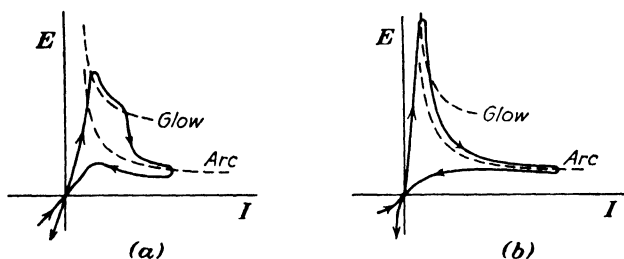


FIG. 11-17.—A.c. arc characteristics with nonthermionic electrodes.

in the neighborhood of the new cathode is never high enough to allow the arc to be maintained with the voltages that will maintain a carbon arc. Ordinarily, the discharge must start as a glow and will change over to an arc only after the current has started to increase. However, with some short heavy-current arcs, especially those having iron or coated-iron electrodes such as are used for a.c. welding, for example, the ion density may be high enough so that reignition occurs with only slightly higher voltages than are required for a carbon arc. Because of their special electrode coating, these arcs are probably partly maintained by thermionic emission. In the case of long heavy-current arcs and short low-current arcs, deionization occurs so rapidly during the period of reversal that the voltage must always rise to the glow voltage, at least, before reignition occurs. When the voltage exceeds the glow voltage, a stable discharge can be formed which can suddenly change to an arc when the current reaches a high enough value; about 0.1 amp. Although with both long heavy-current arcs and short low-current arcs with nonthermionic electrodes the voltage always rises higher than the minimum glow

voltage before reignition occurs, it is only rarely that a definite glow discharge is visible. In most cases, breakdown occurs directly to an arc as it would in a spark gap. The breakdown voltage, however, is lower than it would be in a spark gap with the same spacing because of the presence of the ionization from the preceding conducting period. Figure 11-17 shows schematically the type of volt-ampere curve that is obtained with these arcs. In both sketches the stable arc and glow characteristics are shown dotted. The glow characteristic differs very little from the arc characteristic except in being about 200 or 300 volts higher because of the large cathode fall. In Fig. 11-17a the

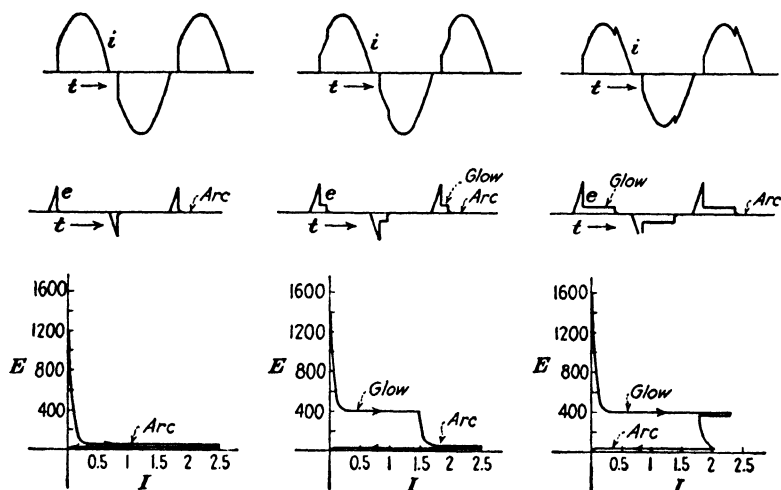


FIG. 11-18.

FIG. 11-19.

FIG. 11-20.

FIG. 11-18.—Cold-cathode arc characteristics. Breakdown to arc directly.

FIG. 11-19.—Cold-cathode arc characteristics. Breakdown to glow followed by breakdown to arc.

FIG. 11-20.—Cold-cathode arc characteristics. Breakdown to glow followed by breakdown to arc late in cycle.

solid curve represents the type of volt-ampere curve that is obtained when a glow discharge forms, and in Fig. 11-17b the solid curve represents the type of curve obtained when a stable glow does not form. The mechanism of arc reignition is essentially the same in each case. On reversal of voltage, there is no cathode spot on the new cathode and the voltage rises above the glow curve. At first the whole cathode collects positive ions as in any glow discharge; very soon after the current exceeds about 0.1 amp., however, the discharge changes quite suddenly

to an arc. When this occurs, the cathode positive-ion current converges to a small spot having a high current density and at the same time the voltage necessary to maintain the discharge drops to a value corresponding to the arc voltage. As the current is decreasing the voltage may sometimes suddenly tend to rise toward the glow characteristic when the current becomes too small to maintain a cathode spot. This sudden increase in voltage with decreasing current is observed only rarely. Typical voltage-time, current-time, and volt-ampere curves of short a.c. arcs between copper electrodes are shown in Figs. 11-18, 11-19, and 11-20. Figure 11-18 shows the characteristics obtained when no visible glow appears. The reignition voltage, however, is always greater than the normal glow voltage. Figure 11-19 shows the characteristics when a glow appears for a short time, and Fig. 11-20 shows the characteristics when the glow lasts most of the half cycle. In the case of the volt-ampere curves, only the first quadrant is shown, for the characteristic in the third quadrant has exactly the same shape. The glow discharge is, of course, characteristic of cold electrodes only. With electrodes of tungsten or carbon, reignition depends simply on getting the electrodes hot enough to become good thermionic emitters; and whenever this condition is reached, the reignition voltage needs to be only slightly higher than the arc voltage.

11-8. Extinction of Arcs.—The fact that there is a definite time lag between changes in current and ion density in an arc is very important as far as arc extinction is concerned. Although it might seem at first glance that this time lag could be an undesirable characteristic of an arc in that it delays the instant of circuit interruption, actually quite the opposite is true; for during the time that deionization is occurring, the energy stored in the inductances of the circuit can be dissipated and large over-voltages can be prevented. In the case of a.c. circuits, of course, the energy stored in the magnetic field is zero when the current is zero; if the circuit could be interrupted at that instant, no over-voltages would occur, but it would take exceedingly close timing to open the circuit just when the current is passing through zero. In most practical a.c. and all d.c. circuits, there is no way of synchronizing the operation of a switch with the instant of zero current; and so we can almost state as a general rule that when a circuit is interrupted some provision must be made for the

dissipation of the energy stored in the inductances of the circuit. An arc, then, is far from being undesirable in a circuit-interrupting device and is, in fact, the means of making a successful circuit breaker or switch, for successful circuit interruption depends not on preventing arc formation but on extinguishing the arc at the earliest practical moment consistent with the allowable circuit voltages.¹

The arc in itself could give rise to high overvoltages if the static characteristic were followed exactly as the current is reduced to zero. Our discussion of dynamic characteristics has shown, however, that wide departures from the static characteristic occur in the neighborhood of zero current and that while the arc voltage and arc current pass through zero together the ion density and gas temperature remain high and the arc space retains a definite conductivity when the current passes through zero. Because of this, it is possible, as we have seen, to have reignition of the arc at potentials considerably below those necessary for sparking in an un-ionized gap of the same length.

In a d.c. circuit the only way that an arc can be extinguished is by increasing the arc drop to a point where the arc is unstable. Practically, this is accomplished either by increasing the arc length or by increasing the rate of deionization along the length of the arc. Usually, both methods are used. In d.c. contactors and low-voltage circuit breakers, the arc is often drawn out by the separating contacts and at the same time the current to be interrupted is passed through a coil of a few turns to produce a transverse magnetic field which blows the arc up a chute and greatly increases its length while at the same time increasing the rate of deionization by the proximity of the chute walls. In high-voltage d.c. breakers the arc drop is further increased by immersing the arc in oil, projecting a blast of oil or compressed air against the arc, or by forcing the arc against a stack of fiber or metal plates. In all these methods, the principal object is to provide a large quantity of cool un-ionized vapor close to the arc core or to force the arc against cool deionizing surfaces so that the loss in ionization by diffusion and the loss in heat by thermal

¹ For a discussion of the relationship between circuit characteristics and arc extinction, the reader is referred to J. Slepian, *J. Franklin Inst.*, **214**, 413, 1932, and *Trans. A.I.E.E.*, **47**, 1398, 1928.

conduction will increase the voltage necessary to maintain the arc to a higher value than the circuit can supply.

In some d.c. breakers,¹ arc extinction is facilitated by breaking the arc up into many small arcs in series. In these breakers the arc is forced by its own or an auxiliary magnetic field into a V-shaped slot in a series of insulated iron and copper plates set perpendicular to the axis of the arc. As the arc is forced into the bottom of the V, not only do the energy losses by diffusion and thermal conduction increase greatly, but the current density in the arc core also increases greatly. The result of this is that the metal plates quickly charge up to a considerably negative potential with respect to the arc—because of the greater mobility of electrons—and, with the high current density already adjacent to the metal, a cathode spot forms. With many arcs in series, then, instability in any one will mean that all will be extinguished.

Very often, trick circuits are used to extinguish d.c. arcs, especially in small breakers and switches. These usually depend upon the generation of oscillations or pulsations in the arc current large enough so that the arc current actually tends to reverse momentarily. Arc extinction in these cases is essentially the same as in any a.c. circuit.

In a.c. circuits, because of the dynamic characteristics of an arc, the arc current and voltage always pass through zero together and if extinction occurs, it is always following this transition. The type of voltage-current characteristic occurring when an a.c. arc is extinguished is shown by curve *a*, Fig. 11-21. As shown here, the current attains a quite definite negative value before extinction occurs. If the initial current is larger or if the voltage rises rapidly after current zero as in a highly inductive circuit, extinction may not occur (curve *b*). To recapitulate the processes that are going on in the arc, we can say that while the current is increasing a high density of ionization is being built up in the arc positive column. Then, as the current decreases, this large density is no longer necessary and the ions start to

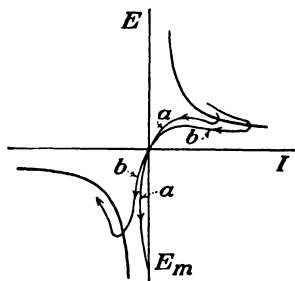


FIG. 11-21.

¹ SLEPIAN, J., *Trans. A.I.E.E.*, **48**, 523, 1929.

diffuse away. If the current decreases rapidly, not a great deal of ionization will be lost before the circuit voltage reverses. Then, as the circuit voltage increases, the arc voltage will increase and the direction of motion of the ions will change. If too many ions have not disappeared, the arc will be reestablished with the available voltage. However, if the current has decreased less rapidly or has not had so high a maximum value, then, when the voltage reverses, there will not be sufficient ionization present to reestablish the arc, even with the maximum available voltage E_m . Essentially, arc extinction depends both on the rate of deionization during the reversal of current and the natural period of the circuit itself. If deionization occurs rapidly, if the natural period of the circuit is long, or if the circuit is of nearly pure resistance, reignition will not occur and the arc will be extinguished. On the other hand, if deionization is slow, if the natural period of the circuit is low, or if the circuit is highly inductive, reignition may occur cycle after cycle until some other effect, such as an increase in arc length, increases the loss by deionization during the transition and makes extinction possible.

We saw above in Sec. 11-3, Fig. 11-8, that the total loss in energy from a free arc is a minimum for a particular arc radius. For smaller values of r , the loss by diffusion increased rapidly; for larger values of r , the loss by thermal conduction caused an increase in the total loss. Since with constant current the energy input to the arc must be equal to the energy lost, it follows that as the arc radius is restricted more and more the voltage necessary to maintain the arc increases rapidly. Similarly, the voltage necessary to maintain the arc would be increased if the arc could be made to occupy larger and larger volumes. The change in arc voltage with arc radius is of importance in a.c. arc extinction, for when the arc is confined to a very small hole the high deionization rate raises the reignition voltage. Furthermore, the value of the maximum current is no longer important in determining the reignition voltage because the diameter of the arc is fixed by the diameter of the hole and consequently the rate of loss of ionization during current reversal is determined by the size of the hole rather than by the magnitude of the current. When the arc is drawn in a small fiber tube or slot, the vaporization of the tube walls contributes to raising the reignition voltage still further. In this case, the value of the

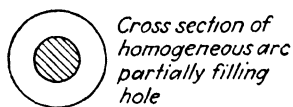
maximum current has a large effect, for it determines the amount of vapor that will be generated at the tube walls. Since the arc completely fills the tube, the rate of deionization by diffusion is already high; but the continual introduction of fresh gas creates a turbulent motion of the vapor in the arc and brings the ions to the walls even sooner. The reignition voltage and the arc voltage can thus be raised by a factor of about 5.

If, however, the hole diameter is slightly larger than the natural diameter of the arc, a large gas evolution from the tube walls can have the effect of raising the arc drop to 10 or 20 times its normal value. The indications are that in this case the effect of the turbulence is to increase the arc cross section until it fills the hole. In a particular case¹ a 2,500-amp. arc in a 1.58-cm. hole in soapstone required a reignition voltage of 40 volts/cm. This is essentially the same as for a free arc in air. However, when the arc was drawn in the same diameter hole in horn fiber, the reignition voltage was raised to 400 volts/cm. Calculations show that the increased reignition voltage cannot be due simply to the heating of the gas that is evolved, for the maximum amount that this could increase the voltage is about 10 or 20 volts/cm. of arc length. Neither can we say that the increased voltage is due to the fact that energy is necessary to produce decomposition of the fiber. The increased voltage must be due to the turbulence in the arc produced by the introduction of a large quantity of un-ionized gas and the resultant increased rate of deionization. Still, the question is not completely answered, for arc extinction occurs more rapidly than can be explained simply on the basis of turbulent gas flow moving ions to the walls as in the case of an arc in a *small* fiber tube. In the time necessary for arc extinction, 10 to 100 microsec., the gas would have a chance to move only a fraction of a centimeter even with the high gas velocities observed (about 10^3 m./sec.). The only satisfactory explanation seems to be that the positive column of the arc actually changes its character. This is the explanation offered by Slepian.² According to his theory, the turbulence, besides causing the arc to assume a larger cross section than it normally would have, produces a nonuniform ion distribution which has the effect of breaking the arc up into many small fila-

¹ SLEPIAN, J., *Trans. A.I.E.E.*, **49**, 426, 1930.

² *J. Franklin Inst.*, **214**, 413, 1932; general reference 11-1, p. 178.

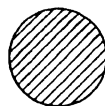
ments of highly ionized gas surrounded by rapidly moving gas of much lower ionic density with cold gas constantly being stirred into the arc space. In Fig. 11-22 the situation is illustrated schematically, the arrows indicating the direction of flow



Cross section of homogeneous arc partially filling hole



Cross section of turbulent filamentary arc



Cross section of uniform arc completely filling hole

FIG. 11-22.

of gas. Because of the falling volt-ampere characteristic of the arc, two arcs in parallel will have a higher drop than a single arc of the same total current. Similarly, with many filamentary arcs in parallel, the drop will be much higher than for a single homogeneous arc of the same total current. The filamentary character of the arcs, the enlarged cross section, and the increased rate of diffusion all contribute to raising the arc drop to 10 or 20 times its normal value. As shown in Fig. 11-23, the volt-ampere

curve for the filamentary type of arc will be above that for a uniform nonturbulent arc. Since all ion generation occurs in the small filaments, when the current is near zero, ion generation will be a minimum and the effect of the turbulence will be to remove

all traces of the filaments, a more or less uniform state of low ionization being thus produced. The reignition voltage for this type of ionization will be even higher than for the filamentary arcs, for there is no region in which a density of ionization occurs comparable with that in the core of an arc. We may represent the voltage necessary for reignition under these circumstances by still a third volt-ampere curve lying above that for the filamentary arc.¹ Thus, if reignition is to occur, the actual

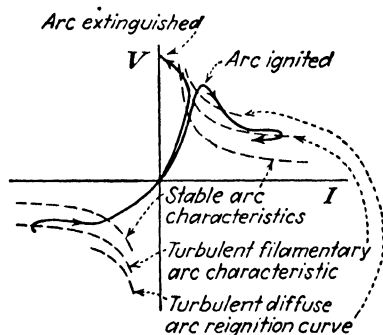


FIG. 11-23.—Dynamic characteristics of a turbulent arc.

¹ There is no question in this case of the dynamic curve rising above the glow characteristic before reignition occurs, for the filamentary-arc characteristic itself is usually above the glow characteristic.

volt-ampere characteristic of a turbulent arc must rise higher than the reignition voltage indicated by this curve. Reignition cannot occur unless this is true even though the actual arc characteristic crosses the characteristic for the filamentary type of arc. In this respect, turbulent arcs are quite different from nonturbulent arcs; and, because of this distinction, it is possible to have circuit breakers that will open circuits of thousands of amperes. Experiments show that it is actually the turbulence in the arc region which is effective in raising the reignition voltage, not the introduction of fresh gas.

11-9. Circuit Interrupters.—The greatest application of the principles and characteristics of a.c. arcs is found in the design of high-power circuit breakers. Although the interruption of a d.c. arc is more difficult than the interruption of an a.c. arc, since it usually requires raising the arc drop to a higher value than can be maintained by the power source, high-power d.c. arcs are at present encountered at only moderate voltages and a.c. breakers can be adapted for the purpose. Alternating current, on the other hand, is used for most present-day power transmission; and a.c. breakers must be able to interrupt currents of thousands of amperes at voltages as high as 287 kv. Fortunately, as the discussion in the previous sections has shown, interruption of an a.c. arc is not so much a problem of extinguishing an arc as of preventing its reignition. In a d.c. circuit the arc current has to be reduced to zero by the action of the circuit breaker itself, but in an a.c. circuit the current automatically goes to zero 120 times a second. The problem then is simply to prevent the arc from restriking. However, this is not so easy as it seems, especially in high-voltage circuits.

In the design of circuit breakers the primary objective is to produce a breaker that will open the circuit in the shortest possible time. Fast operation is necessary not only to prevent damage to the apparatus caused by excessive currents but also to prevent burning of the conductors at the point of short circuit, to reduce the magnitude and the duration of voltage dips on the system, and to prevent instability in the system as a whole. This means that ordinarily the breaker must open the circuit within less than 8 cycles after the fault occurs. The tendency in present-day design is to try to reduce this to even shorter times; actually, most breakers now operate within 5 cycles.

In the case of the Boulder Dam line, operation time has been reduced to less than 3 cycles. This is exceptionally short; although breakers have been made to clear a fault within 1 cycle, most operators are satisfied with 3- to 4-cycle operation. This is over-all time from the instant the fault occurs until it is cleared and is distributed approximately as follows: $1\frac{1}{2}$ to 2 cycles to energize the tripping circuit of the breaker through the auxiliary relays, $\frac{3}{4}$ to $1\frac{1}{4}$ cycles to set the breaker contacts in motion, and $\frac{1}{2}$ cycle or less to interrupt the arc. In practically all high-voltage breakers, the duration of the arc has now been limited to $\frac{1}{2}$ cycle so that the circuit can be opened at the first reversal of current after the contacts start to move. Shorter arcing times than this cannot, of course, be obtained. At present, circuit-breaker development is along the line of increasing the current and voltage capacity of the breaker.

The operation of any breaker is very strongly dependent on the type of circuit in which it is used. For example, a certain conventional breaker was able to open a 10,000-volt (peak) circuit with a contact separation of 1 in. when it was used in a pure resistance circuit; but, in a highly inductive circuit, it opened less than 1,000 volts/in. of contact separation.¹ The reason for this is evident when we remember that circuit interruption or reignition really depends on the outcome of a race between deionizing agents and circuit voltage. If the circuit voltage rises rapidly enough after reversal, sufficient deionization cannot occur to prevent reignition of the arc. On the other hand, ample time for deionization occurs when the circuit voltage rises slowly, as in a resistance circuit, and the arc is extinguished.

In a practical circuit the arc voltage never rises instantaneously to a high value when the arc is extinguished at a current zero, as it would in the circuit of Fig. 11-24, for all circuits possess a certain amount of distributed capacitance. Conditions are more nearly like those indicated in Fig. 11-25 where C is the distributed capacitance between turns of the reactance coil L . Because of this distributed capacitance the counter-voltage of the inductance does not appear instantly across the breaker contacts when the current stops but builds up gradually in one-quarter of the period of natural oscillation of the LC circuit as shown on the arc-voltage curve. We also find in the

¹ BAKER, B. P., and H. M. WILCOX, *Trans. A.I.E.E.*, **49**, 431, 1930.

practical case that the maximum voltage across the breaker contacts is not just the maximum circuit voltage. Because of the oscillatory nature of the circuit the breaker voltage may reach a value of twice the peak circuit voltage on the first oscillation. In practice, it is found that the frequency of these oscillations

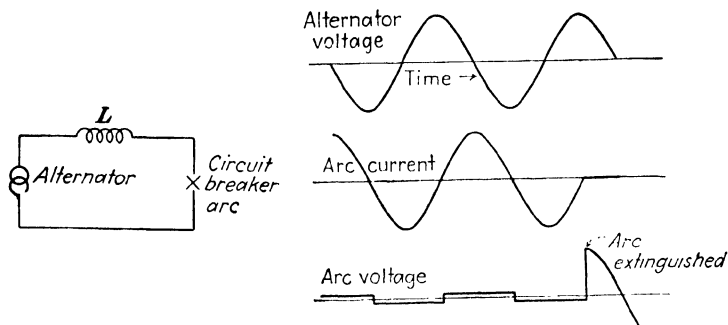


FIG. 11-24.—Ideal inductive circuit.

may vary from about 100,000 cycles/sec. in a simple circuit with a current-limiting reactor to a few hundred cycles per second in a circuit containing a long transmission line.

Experiments have shown that there are two ways in which a high reignition voltage can be obtained within 1 or 2 microsec.

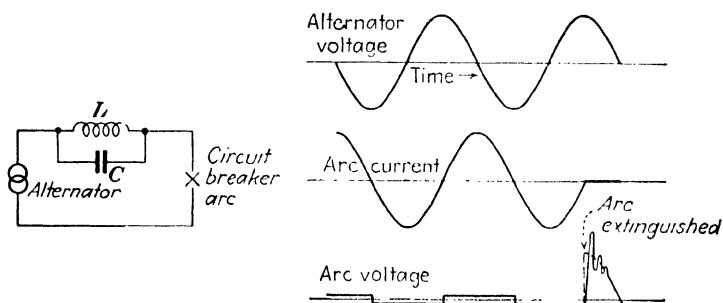


FIG. 11-25.—Actual inductive circuit.

after current zero. One method is simply based on the fact that when reignition occurs between cold electrodes the discharge always starts as a glow although it may subsequently change to an arc. Because of this, the circuit voltage must reach a value at least equal to the normal cathode fall in the glow before the discharge can restrike. In air between copper electrodes, this

is about 235 volts. By using many short arcs in series the reignition voltage can thus be raised to a high value. This is the principle employed in the Deion air circuit breaker.¹ The other method depends on drawing the arc out to a length of several inches and then, by creating turbulence in the arc core,

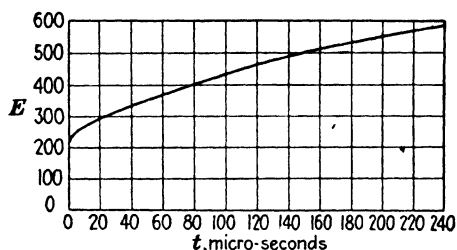


FIG. 11-26.—Reignition characteristic for 1.6-mm arc between copper electrodes (short arc). (*J. Slepian, Trans. A.I.E.E.*, **49**, 421, 1930.)

preventing reignition in the manner described in the previous section. This is the most generally used method in oil circuit breakers. In either case the reignition voltage increases gradually with time after the current reaches zero because of the continual loss in ionization before reignition occurs. With

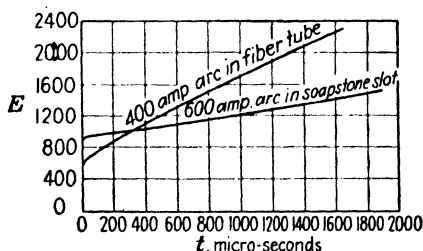


FIG. 11-27.—Reignition characteristics for turbulent arcs (long arcs). (*T. E. Browne, Trans. A.I.E.E.*, **51**, 185, 1932.)

either of these methods, then, the arc space acquires an initial dielectric strength of several hundred volts; but this is soon increased by the deionizing agents acting, and unless the natural period of oscillation of the circuit is very high the arc space soon acquires the ability to withstand a very high voltage. These characteristics are indicated in Figs. 11-26 and 11-27 where the abscissas represent time after the first current zero and the

¹ SLEPIAN, J., *Trans. A.I.E.E.*, **48**, 523, 1929.

ordinate is the minimum voltage necessary for reignition. Two curves are shown for the turbulent arc. One of these was obtained by rotating a 3.8-cm.-long arc in an annular slot 1.9 cm. wide in soapstone. In this case a linear speed of 68 m./sec. was produced by a d.c. radial field of 0.034 weber/sq. m. and was sufficient to produce turbulence in the arc. In the other curve, turbulence was produced by a 400-amp. arc, 7.6 cm. long, in a 1.6-cm.-diameter fiber tube. In this case the velocity of the gas through the tube was calculated to be approximately the same as in the soapstone slot. The curves do not coincide, however, because the arcs are of different lengths and the maximum values of current are different. The curves indicate, however, that in a very short time following current zero the arc reignition voltage rises to a high value.

The problem of arc extinction in a high-voltage fuse is essentially the same as that in any circuit breaker with the exception that very high speed is usually not necessary. Fuses are made, however, that interrupt arc currents up to 20,000 amp. with circuit voltages of 13,800 volts r.m.s. within 1 or 2 half cycles after the fuse link melts. These fuses invariably make use of arc turbulence for interruption although the methods employed to produce the turbulence vary. In some cases a piston is arranged to increase the gas pressure and expel gas through a vent; in others the arc is drawn in a small fiber tube which partly decomposes in the heat of the arc and expels large quantities of gas, turbulence being created in the arc space as in the first case. In still other fuses, the arc is drawn in substances such as boric acid that decompose readily under the action of heat and liberate large quantities of water vapor to produce the turbulence. In all cases the reignition characteristics obtained are similar to those indicated in Fig. 11-27 and distinctly correspond to a turbulent-arc condition.

Circuit breakers can be divided roughly into three general classes. (1) High-voltage breakers for 138-kv. voltages and higher. The currents that these breakers must interrupt are usually less than 7,000 amp. per phase, and the function of the breaker is principally to preserve system stability. Thus, these breakers must be able to interrupt very high voltage arcs and restore voltage to the system as soon as possible. (2) Breakers in the 15-kv. class. These are essentially power-plant breakers.

Although this voltage is relatively easy to handle with an ordinary breaker and system disturbances constitute a relatively minor operating problem, the short-circuit current that the breaker must clear may be as high as 100,000 amp. Furthermore, the breaker must operate time after time without maintenance or inspection and without throwing oil or emitting

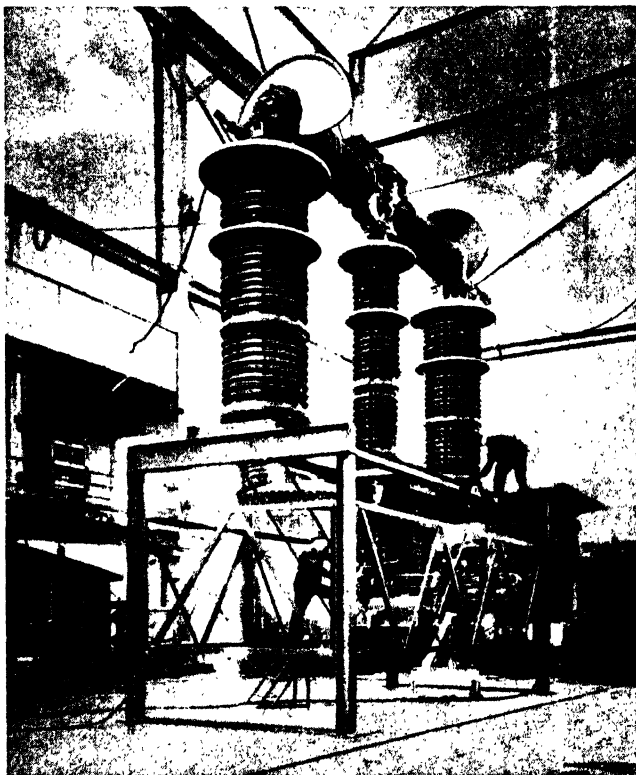


FIG. 11-28.—287-kv. Boulder Dam circuit breaker, oil-blast type. (Courtesy of General Electric Co. and Bureau of Power and Light of the City of Los Angeles.)

flames which could start fires. Thus, the principal problem is to design a breaker that will handle both a fairly high voltage and a high current. (3) Low-voltage breakers, breakers for 600 volts or less. Here the voltage problem is relatively simple, but the current that must be interrupted may be as high as 200,000 amp. These are essentially current interrupters.

High-voltage breakers all make use of turbulence for arc extinction although the methods employed to produce the tur-

bulence differ in various designs. Since breakers for 138-kv. voltages and higher are all oil breakers, the turbulence arises from the introduction of oil vapor into the arc stream. In general, there are only two ways that this can be done. One is by moving oil into the arc, and the other is by moving the arc

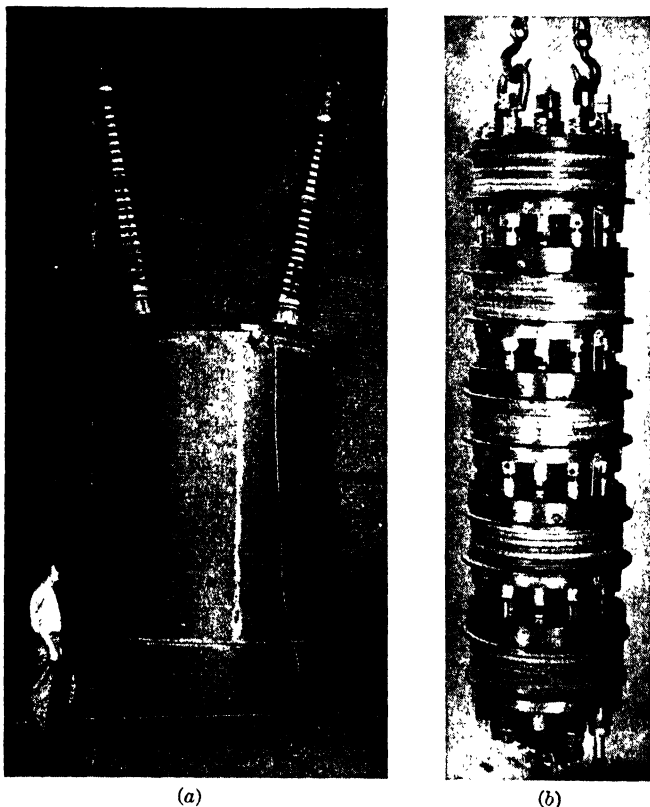


FIG. 11-29.—287-kv. Boulder Dam circuit breaker, Deion type. (a) complete breaker; (b) multibreak grid structure from one pole of the breaker. (Courtesy of Westinghouse Electric and Manufacturing Co.)

into the oil. Actually, the oil can be moved parallel or perpendicular to the arc core, and combinations of the two methods can be used, several different designs being thus possible.

In one type used on the 287-kv. Boulder Dam line¹ (see Fig. 11-28) a transverse blast of oil not only stretches the arc out to several times the electrode separation but by its high velocity

¹ PRINCE, D. C., *Trans. A.I.E.E.*, **54**, 366, 1935.

sets all the oil vapor, in which the arc actually exists, in very turbulent motion and by increasing the rate of ion diffusion greatly increases the reignition voltage after current zero. In another type of breaker used on the Boulder Dam line¹ (see Fig. 11-29) the arc is drawn across a series of oil-saturated fiber plates; then by means of a magnetic field—the magnetic field due to the arc current itself assisted by iron plates imbedded in the fiber—the arc is forced into a narrow slot in the fiber plates. Here the turbulence is produced by oil vapor projected directly into the arc stream from the oil-saturated plates, or “grids,” by the heat of the arc itself. These breakers are both rated at 2.5 million kilovolt-amperes at 287 kv., and both operate in less than 3 cycles. Because there is a limit to the speed at which arc contacts can be moved, these breakers, as well as all other high-voltage breakers, gain operating speed by using a multibreak construction. That is, in the first breaker 8 arcs and in the second 10 arcs are drawn in series as soon as the breaker starts to open. A more recent design² for a 138-kv. breaker employs a blast of oil parallel to the arc at the same time that the arc is drawn in a narrow slot or well between two oil-saturated corrugated walls of insulating material (see Fig. 11-30). In this case the turbulence arises not only from a transverse blast of oil vapor driven out of the oil-saturated walls of the chamber by the heat of the arc but also from a longitudinal blast of oil vapor due to the vaporization of the oil flowing parallel to the arc. In tests, this breaker interrupted 5,900 amp. at 124 kv., or 2.2 million kilovolt-amperes on a three-phase basis with a total short-circuit time of less than 4 cycles.

At 15 kv., essentially the same principles of operation are employed in the design of oil circuit breakers of higher current-rupturing capacity;³ but perhaps the design of greatest theoretical interest is the 15-kv. Deion air breaker,⁴ for it is an outstanding example of the application of arc-reignition characteristics to the design of a practical breaker. Since the initial recovery

¹ WILCOX, H. M., and W. M. LEEDS, *Trans. A.I.E.E.*, **55**, 626, 1936.

² BAKER, B. P., *Trans. A.I.E.E.*, **60**, 440, 1941.

³ PRINCE, D. C., and W. F. SKEATS, *Trans. A.I.E.E.*, **50**, 506, 1931. SPURCK, R. M., *Trans. A.I.E.E.*, **51**, 171, 1932.

⁴ SLEPIAN, J., *Trans. A.I.E.E.*, **48**, 523, 1929. DICKINSON, R. C., and B. P. BAKER, *Trans. A.I.E.E.*, **48**, 528, 1929. DICKINSON, R. C., *Trans. A.I.E.E.*, **58**, 421, 1939.

voltage of any arc between cold electrodes is about 235 volts (Fig. 11-26), this breaker utilizes this recovery voltage by splitting the arc up into nearly 100 short arcs in series. Practically, this is accomplished by using a transverse magnetic field to blow the

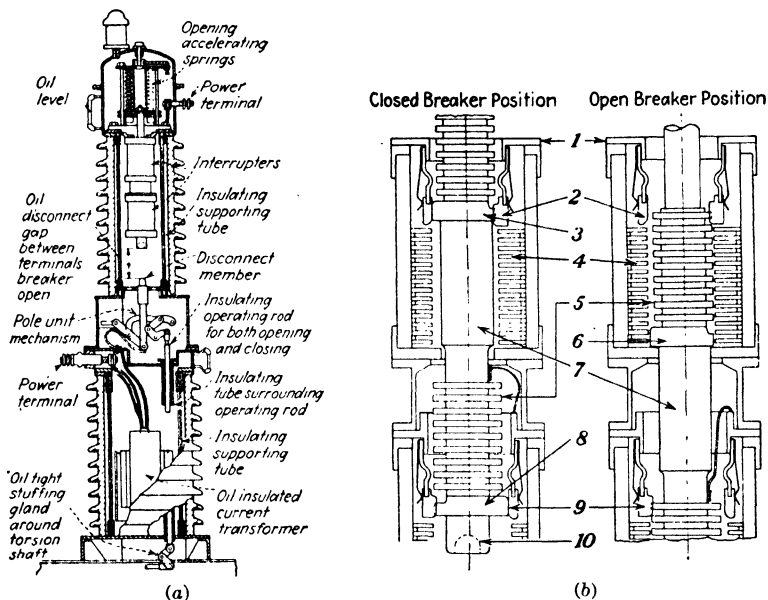


FIG. 11-30.—138-kv oil-poor circuit breaker, (a) cross section of assembled breaker; (b) sectional view of interrupters.

- | | |
|--|---|
| (1) Upper terminal connects here; | (7) Metal cylinder carrying upper contact. |
| (2) Upper fixed contacts. | (8) Lower movable contact (closed position). |
| (3) Upper movable contact (closed position). | (9) Lower fixed contacts. |
| (4) Corrugated walls (insulation). | (10) Point of contact with disconnect member. |
| (5) Corrugated plungers (insulation). | |
| (6) Upper movable contact (open position). | |

(Courtesy of Westinghouse Electric and Manufacturing Co.)

arc up into a series of V-shaped slots in insulated copper plates spaced $\frac{1}{8}$ in. apart (see Fig. 11-31). Then, because of the high current density close to the insulated copper plate, a series of new cathode spots forms on the successive plates and the arc is split up into many arcs in series. To prevent the heat generated in the arc from melting the plates before the current goes to zero, another magnetic field is used to rotate the arc rapidly around a circular path on the copper. Then, when the current

goes to zero, since all parts of the arc are close to relatively cool deionizing surfaces and at least 235 volts per gap is needed anyway to reignite the arc, the circuit is interrupted. These breakers have been used to interrupt currents up to 3,500 amp. in a 12-kv. circuit with operation times of less than 4.5 cycles. They have also been used for d.c. circuit interruption when the circuit voltage did not exceed 0.2 of the limiting a.c. voltage.

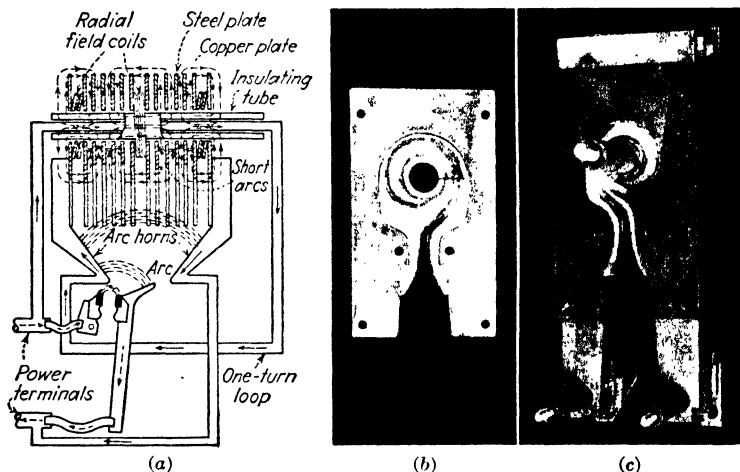


FIG. 11-31.—Type U Deion air circuit breaker. (a) Typical arrangement of the component parts of a Deion breaker, showing the method of forcing the arc up into the deionizing chamber by means of a magnetic field; (b) typical deionizing plate consisting of a copper insert in a steel plate (note the path of the arc on the copper); (c) partially assembled deionizing chamber of a 2,500-volt breaker (Type 25-U-25), showing deionizing plates, insulating spacers, and conductor studs for the blow-in magnetic field. (Courtesy of Westinghouse Electric and Manufacturing Co.)

The low-voltage high-current breakers are almost all air breakers. These usually depend upon convection and a transverse blowout magnetic field to force the arc up against surfaces on which recombination can easily occur during the time that the current is passing through zero, in addition to lengthening the arc enough so that the available circuit voltage cannot maintain it. These breakers also depend on turbulence for arc extinction although to a lesser extent than the high-voltage breakers where very high voltage recovery rates are observed. One design of low-voltage breaker¹ interrupted peak currents

¹ LUDWIG, L. R., and G. G. GRISSINGER, *Trans. A.I.E.E.*, **58**, 414, 1939.

of 52,000 amp. at 575 volts or 204,000 amp. at 288 volts when the three contacts were connected in parallel on a single-phase circuit; it also interrupted r.m.s. currents per phase of 38,200 amp. on a 600-volt three-phase circuit and d.c. currents of 62,400 amp. at 750 volts.

We cannot leave this subject without saying a word about the contradictory theories that have been offered to explain the successful operation of the various types of oil circuit breakers.^{1,2} Actually, of course, because of the enormously high temperature in the positive column of a high-pressure arc, the arc always plays in a bubble of gas or oil vapor; because of the definite viscosity of the oil and the vapor, this bubble cannot be completely dissipated or removed in the short time involved in voltage reversal, as has been proposed. Furthermore, forcing the arc against a wedge or splinter of insulating material (solid or liquid) cannot result in cutting the arc in two and thus interrupting it, for at the surface of the splinter the vapor velocity must be zero for hydrodynamic reasons. Thus, all that occurs is that the arc is bent around the splinter. The reason why extinction does occur so easily at current zero is not that the arc is "cut" but that a high-velocity vapor stream projected into the arc either by a gas blast or by forcing the arc against an oil-soaked splinter introduces so much turbulent motion into the arc stream that the loss in ionization by diffusion is increased by a factor of 100 or more and a very high reignition voltage is necessary to reestablish the arc. High-speed photographs of an arc during the period of current reversal indicate that, when a vapor stream is projected into the arc, it does not simply displace the ionized gas but actually mixes with it and breaks the uniform arc stream up into splotches of highly ionized and radiant gas by its turbulent motion. Slepian has recently reviewed the arguments against this so-called "displacement" theory of arc extinction by cutting the arc with an insulating medium and has shown by an analogy with a problem in aerodynamics that turbulent motion can increase the loss in ionization by diffusion by as much as 100 or 1,000 times and that this is the basic explanation for high-speed arc interruption under oil.

¹ PRINCE and SKEATS, *op. cit.* PRINCE, D. C., *Trans. A.I.E.E.*, **51**, 166, 1932.

² SLEPIAN, J., *Trans. A.I.E.E.*, **60**, 162, 1941.

11-10. Mercury-arc Rectifiers.—The fact that a cathode spot is necessary for the maintenance of a low-voltage arc between cold electrodes makes it possible to utilize the arc for the purpose of converting alternating to direct current. If a cathode spot is maintained on the surface of one of a pair of electrodes in a low-pressure discharge tube and an alternating potential is applied between the electrodes, current will flow through the gas and the external circuit whenever the electrode with the cathode spot is negative with respect to the other by about 10 volts. When it is positive with respect to the other electrode, only a very small positive-ion current will flow until the tube voltage attains a

value high enough for the establishment of a self-maintaining discharge. If this voltage is not reached, current will flow in only one direction and rectification will occur.

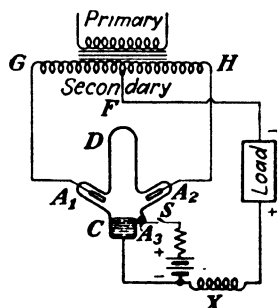


FIG. 11-32.—Single-phase mercury-arc rectifier circuit.

The mercury arc rectifier is a device of this type. It usually consists of a chamber, either metal or glass, evacuated to as high a degree as possible and containing a mercury pool for one electrode and one or more other electrodes of metal or graphite. Figure 11-32 represents diagrammatically a two-anode rectifier and its auxiliary circuit. When the cathode spot is maintained on the mercury surface, the arc forms between it and either A_1 or A_2 , depending on which is positive with respect to C . As we saw in Sec. 11-4, the temperature of the cathode spot is certainly not high enough in a mercury arc to cause thermionic emission, but a temperature of even 250 or 350°C. is sufficient to produce rapid vaporization of the mercury at the pressures that are used. Thus, although the foreign gas pressure in the tube may be only about 10^{-5} mm. Hg, the arc is actually existing in mercury vapor evaporated from the mercury cathode. There is so much vapor leaving the local area of the cathode spot that the reaction of this vaporization of the cathode is to produce a depression in the mercury surface. Enough ionization by collision occurs, then, in the relatively high vapor pressure just over the cathode spot to maintain the discharge.

The advantage of a mercury pool for a cathode over a thermionic cathode in a high-power rectifier is that with the cold

cathode the tube is able to stand very high overloads for short periods of time with no material damage. The current that can be obtained with a thermionic cathode is limited by the temperature of the cathode, and any attempt to overload such a cathode would result in the destruction of the cathode surface. The mercury pool has one disadvantage, however, in that the cathode spot must be maintained if rectification is to persist. This means that in any particular application the circuit must

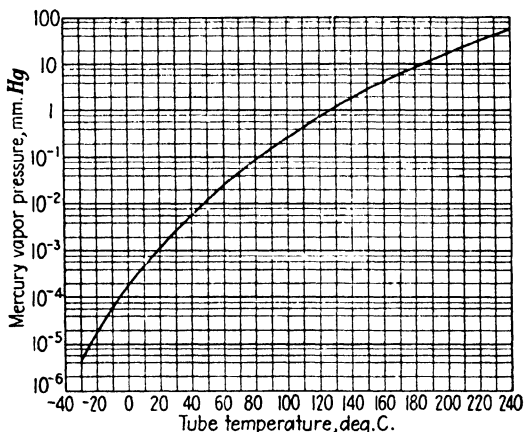


FIG. 11-33.—Dependence of mercury-vapor pressure on temperature. (Data from "Handbook of Chemistry and Physics," Chemical Rubber Publishing Company, 21st ed., p. 1325.)

be so arranged that some current is always flowing into the cathode.

The function of the dome *D*, indicated in Fig. 11-32, is to provide a cool surface on which the mercury vapor can condense and return to the cathode pool. Experiments have shown that if failures, or "arc-backs" (see Sec. 11-11), are to be avoided, the vapor pressure in a rectifier tube should have as low a value as possible. Since the vapor pressure of mercury in a tube depends very strongly on the temperature of the tube walls, as shown in Fig. 11-33, an attempt is made in all rectifiers to have some cool surfaces on which condensation can occur. In practice, it is found that the vapor pressure in any part of a tube does not differ by more than a factor of 2 or 3 from the vapor pressure corresponding to the coolest part. Thus, in large metal-tank rectifiers such as that shown in Fig. 11-35a the tank itself is water-cooled, and water-cooled surfaces are provided inside

the tank on which condensation can occur; in small glass rectifiers, a condensing dome such as is shown in Fig. 11-32 is used.

In initiating the discharge, the cathode spot is formed by the separation of two electrodes between which there is a difference of potential. With the device indicated in Fig. 11-32, this would

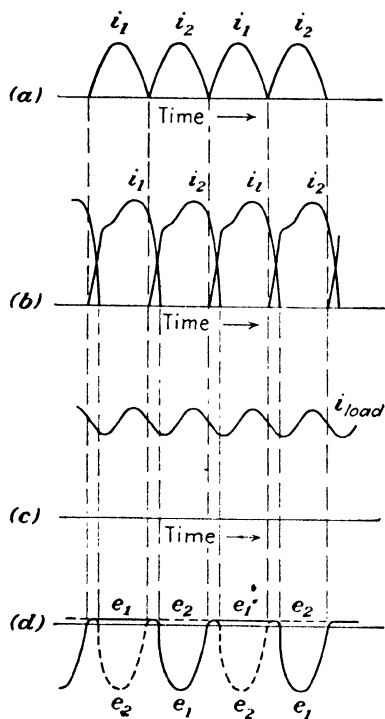


FIG. 11-34.—Voltage and current relations in a single-phase rectifier.

be accomplished by closing the switch S and tilting the tube slightly until the two mercury pools C and A_3 run together. Then, as the mercury bridge is broken, an arc forms. Now, whenever A_2 is positive with respect to C by more than a few volts, current will flow in the circuit $A_2, C, \text{load}, F, H, A_2$. And whenever A_1 is positive with respect to C , current will flow in the circuit $A_1, C, \text{load}, F, G, A_1$. Thus, a pulsating direct current will pass through the load as shown in Fig. 11-34a, and we have changed an alternating into a direct current. In that figure, i_1 represents the current flowing through A_1 and i_2 the current through A_2 . With this arrangement the current falls to zero at the end of each half cycle, and in order to have the

rectification process continue the cathode spot must be maintained by an auxiliary d.c. arc between A_3 and C . It is possible to do away with this auxiliary circuit by introducing a reactance X in series with the load. The function of this is to cause the current to persist in A_2 , for example, until the potential of the transformer has reversed and A_1 has become positive with respect to C . Thus, for a short time in each cycle, both anodes A_1 and A_2 are carrying current (see Fig. 11-34b). The transformer potential varies sinusoidally, but because of X the current through A_1 or A_2 lags behind the voltage and persists even after the

potential of G with respect to F has reversed and become negative. The resultant load current, then, has the form shown in Fig. 11-34c, and the cathode current never reaches zero. Thus the arc is maintained without the aid of the auxiliary circuit. The voltage between anode and cathode does not, of course, reverse when the transformer voltage reverses but depends on the arc current, as shown in Fig. 11-34d. In this figure, e_1 is the voltage between A_1 and C , and e_2 is the voltage between A_2 and C . Because of the characteristics of the arc, the voltage between anode and cathode cannot reverse until the current reverses or stops; thus, during the short interval between the reversal of the transformer voltage and the current zero, the potential for the arc must be supplied by the reactance coil X .

When it is desired to have a rectifier that will deliver a very large current, it is customary to use a steel tank for the containing vessel. These high-power rectifiers are operated from three-phase sources and in order to produce a smoother d.c. voltage are provided with 6 or 12 "anodes" which carry current alternately. Figure 11-35b shows a typical polyphase rectifier in cross section, and Fig. 11-36 shows one type of polyphase-rectifier connection.¹ Here, the power transformer has six secondary windings (two for each phase) arranged so that at least two anodes are always carrying current. That is, each anode carries current for $\frac{1}{3}$ cycle; but every 60 electrical degrees one anode drops its load, and another picks it up. I is an *interphase* transformer that is connected between the neutrals of the two three-phase systems so that an anode on each system will always be carrying current and a keep-alive arc is not necessary to maintain the cathode spot. Without the interphase transformer, the connections would form a simple six-phase circuit, and only one anode would carry current at a time.

Aside from the fact that mercury-arc rectifiers can be used at much higher voltages than any ordinary rotating apparatus, are much quieter in operation, and have no moving parts, the principal advantage of the rectifier is the very high efficiency that can be attained at all loads. Since the power loss in the arc constitutes practically the only loss in the rectifier and since

¹ For other rectifier connections, see "Standard Handbook for Electrical Engineers," 6th ed., p. 998, McGraw-Hill Book Company, Inc., New York, 1933.

the arc drop is essentially independent of the arc current, the losses in a rectifier will vary as the first power of the load current. Thus, the efficiency is approximately constant at all loads, even at very light ones. This property makes a rectifier particularly desirable for applications in which it may run under a light load for a considerable period of time. The efficiency characteristic of a 3,000-kw. 600-volt rectifier unit is contrasted with that of an equivalent rotary converter and motor-generator set in Fig.

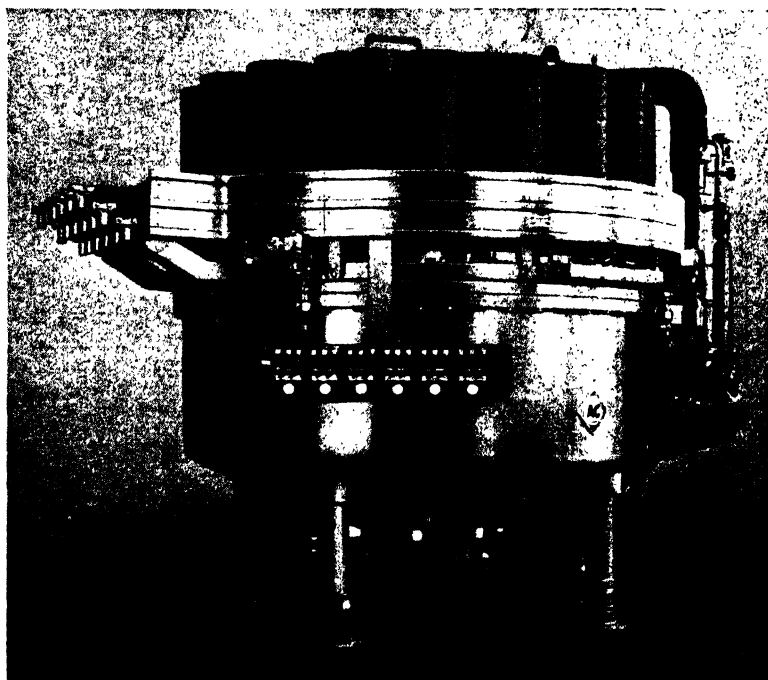
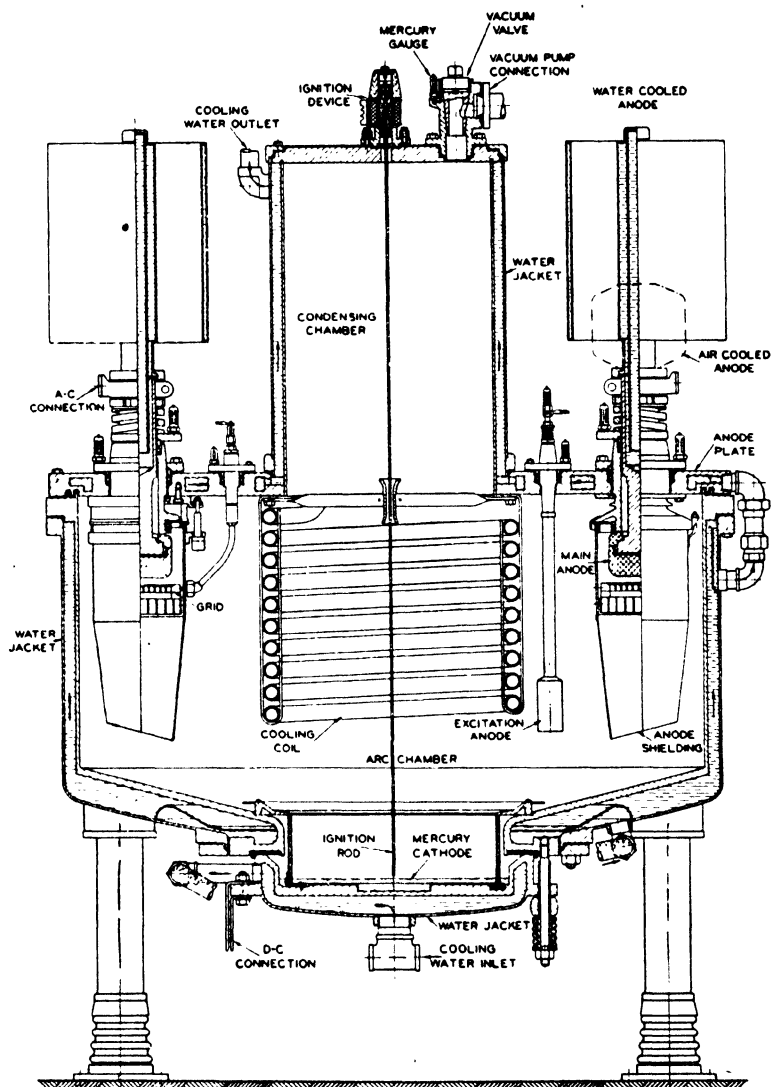


FIG. 11-35a.—Multianode mercury-arc rectifier. (a) 18-anode, grid-controlled mercury-arc rectifier, 3,000 kw., 630 volts; (b) cross section of multianode, grid-controlled, mercury-arc rectifier. (Courtesy of Allis-Chalmers Manufacturing Co.)

11-37. The high efficiency of the rectifier especially at light loads is quite evident from the figure.

Because the arc drop in a rectifier depends only on the arc current and is independent of the circuit voltage, the rectifier has the additional advantage that its efficiency becomes higher as the d.c. output voltage is increased. Thus, if the arc drop in a particular rectifier were 20 volts, the highest efficiency that could



CROSS-SECTION OF MULTI-ANODE MERCURY-ARC-RECTIFIER

FIG. 11-35b.—For descriptive legend see opposite page.

be obtained in a 200-volt circuit would be 90 per cent. But in a 2,000-volt circuit its maximum efficiency would be 99 per cent. With a rotating machine, however, the efficiency would remain essentially the same, regardless of the operating voltage.

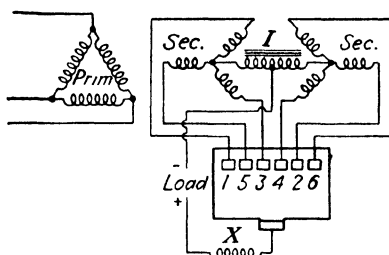


FIG. 11-36.—Double three-phase rectifier connection.

11-11. Arc-backs in Mercury-arc Rectifiers.—Normal operation of a rectifier requires that a cathode spot be maintained on the mercury surface at all times so that whenever one of the other electrodes becomes positive with respect to the mercury pool

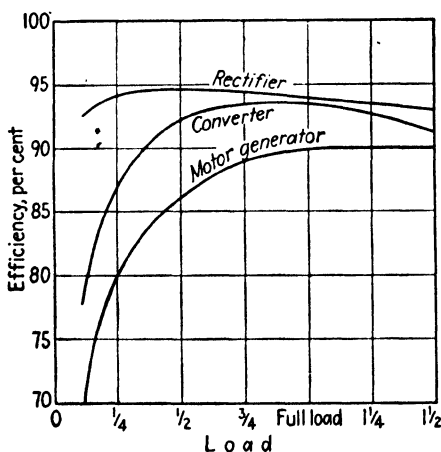


FIG. 11-37.—Over-all efficiencies of 3,000-kw 630-volt mercury-arc rectifier, rotary converter, and motor generator set at 60 cycles/sec. (Courtesy of Allis-Chalmers Manufacturing Co.)

an arc will form to it and a unidirectional current will flow out through the load. Thus, with reference to Fig. 11-38a, current flows either through the transformer winding to A_1 to C through the load and back to the center tap of the transformer

or through A_2 , C , load, and back to the center tap of the transformer, depending on the polarity of the transformer at any instant. In the figure, we have assumed an instant when A_1 is positive with respect to the cathode and current flows along the path of the arrows. Since the full transformer voltage always appears between A_1 and A_2 , the electrode A_2 will always be negative with respect to A_1 while current is flowing through A_1 . A discharge does not ordinarily form between A_1 and A_2 , however; for the whole vessel is evacuated to a low pressure, and there is not a sufficient amount of ionization in the neighborhood of A_2 to support a self-maintaining discharge. To put it another

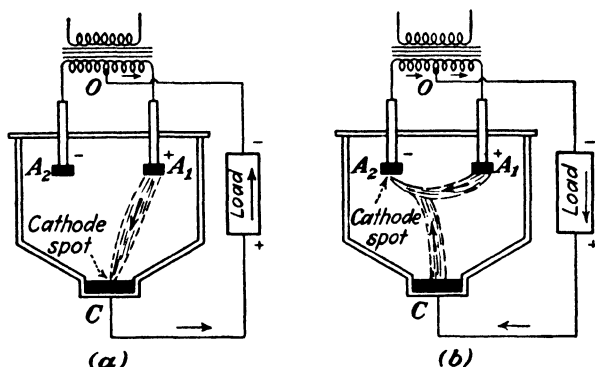


FIG. 11-38.—Conditions in a rectifier, (a) during normal operations; (b) during an arc-back.

way, the pressure is so low in the tube that the available circuit voltage between A_1 and A_2 is well below the sparking potential for the gap. Without the cathode spot on A_1 or A_2 necessary for a low-voltage discharge, current can flow only to the mercury pool. If, now, for some reason a cathode spot should form on A_2 , we have the situation indicated in Fig. 11-38b. Since the full transformer voltage appears between A_1 and A_2 , the arc that forms will result in producing a short circuit on the transformer terminals. The only thing limiting the current in this condition is the impedance of the transformer itself, for the arc drop is small. Furthermore, if the rectifier is operating in parallel with other rectifying equipment or if it is supplying a d.c. motor, the counter e.m.f. of the load can cause a reversal of current through the cathode and further increase the short-circuit current flowing through A_2 . This catastrophe is termed an *arc-back* or *backfire*.

The damage that can be done by an event of this character both to the rectifier and other equipment if the short circuit is allowed to persist is quite evident.

Although in present commercial-rectifier designs for 600-volt service, the occurrence of arc-backs has been reduced to two or three a year, certainly no more frequent than the failures of a rotating machine, arc-backs still furnish a limitation on the use of mercury-arc rectifiers for high-voltage d.c. service. At present, polyphase mercury-arc rectifiers are not being made for voltages much above 3,000 volts direct current.

The cause of these arc-backs has been the subject of much investigation; although many of the effects contributing to the formation of the cathode spot have been determined, it has not been possible, as yet, to eliminate these failures completely. Arc-backs cannot be an ordinary type of sparking phenomena, for it is known that they occur at voltages much lower than those necessary to support a stable glow discharge. Because of the rising character of the sparking potential curve at small values of pd , the sparking potential for the electrode spacings and pressures usually encountered in mercury-arc rectifiers would be of the order of 50 or 100 kv. It is found, though, that arc-backs occur at potentials as low as 5 or 10 kv. and increase in frequency as the potential is raised. The cause of these failures seems to be associated with departures from homogeneity at the negative-electrode surface. For example, experiments have shown¹ that a cathode spot can be produced by mercury drops hitting the electrode or by particles breaking from the electrode, as well as by small patches of insulating material on the electrode surface charging up in the ionized space and breaking down with the formation of a cathode spot. On the other hand, there is considerable evidence² that in the absence of all these effects the cathode spot can still be formed by the emission of a very small amount of gas from the electrode surface. This is entirely possible, for it has been shown that a considerable quantity of gas can be held on or in an electrode surface and released only under the action of positive-ion bombardment. Furthermore, the

¹ SLEPIAN, J., and L. R. LUDWIG, *Trans. A.I.E.E.*, **51**, 92, 1932.

² MAXFIELD, F. A., and G. L. FREDENDALL, *J. App. Phys.*, **9**, 600, 1938. MAXFIELD, F. A., H. R. HEGBAR, and J. R. EATON, *Trans. A.I.E.E.*, **59**, 816, 1940.

amount of gas necessary in a burst to produce or sustain a cathode spot is only about 10^{11} molecules.

Experimentally, it is found that arc-backs are random in their occurrence under any given set of conditions. That is, even though the average frequency of arc-backs may vary with the circumstances of the test, there is no condition under which an arc-back is *sure* to occur. There is a definite chance that an arc-back will occur in a given time interval, but we cannot say surely whether or not it definitely will occur. If the voltage is applied long enough, an arc-back can occur at practically any voltage. Under a given set of conditions, the time intervals between arc-backs will vary over wide limits. Some will be very short and others long, just as the free path of a molecule is long at one time and very short at another. In fact, the distribution of time intervals between arc-backs is exactly the same as the distribution of free paths of a molecule. Slepian and Ludwig¹ have pointed out that, as a consequence of this randomness, care must be used in the interpretation of the results of any test. For example, if it were possible to make a large number of identical rectifiers each with an average arc-back frequency under operating conditions of 1 per month and if each rectifier were tested for 1 month, 37 per cent would not fail at all, and 26 per cent would fail two times or more. It would be entirely erroneous, however, to conclude from this test that 26 per cent of the rectifiers were below standard and 37 per cent above the average; they are all identical, and this distribution is entirely within theoretical expectation, for if all the units were tested for a very long time they would give identical results for the average arc-back frequency.

One of the factors that appears to increase the probability of an arc-back is the magnitude of the positive-ion current collected by the negative electrode. Thus, if the electrodes are close together and unshielded or if the load is highly inductive and a high negative voltage is applied suddenly to one electrode after a conduction period, as in the case illustrated in Fig. 11-34, the negative electrode will attract positive ions from the ionized space in its vicinity and the probability for cathode-spot formation will be high. In general, a high probability for arc-back is related to a large positive-ion current, and anything done to

¹ *Loc. cit.*

a rectifier to reduce this positive-ion current will make it operate more reliably. This explains why it is desirable to place deionizing surfaces such as shields or grids around the electrodes of a high-power rectifier. With these shields or grids, deionization can occur rapidly after a conduction period and few ions can diffuse into the electrode region from the other arcs in the vessel. The fact that a definite time is necessary for the electrode region to become deionized after a conduction period also explains why it is impossible to use mercury-vapor rectifiers for rectification if the frequency is above about 1,000 cycles/sec.

It is also found that the gas pressure has an effect on the arc-back probability, for at high pressures the rate of diffusion of ions out of the electrode region is retarded and the positive-ion current after the current zero persists for a longer time even though its maximum value may be no greater. In addition to this, a high gas pressure seems to aid the development of an arc-back by providing a greater gas density for the production of positive ions. That is, with a high gas density a single high-energy electron will be able to generate more positive ions per unit distance of travel. In order to avoid high vapor densities, some cool surfaces on which condensation can occur are always provided in a mercury-arc rectifier. In a high-power rectifier the whole vessel is water-cooled, and often water-cooled coils are introduced inside the vacuum chamber to provide even more of a cool surface for condensation.

Since arc-back causes are apparently of very short duration—less than 5×10^{-6} sec.—one method of avoiding a rectifier failure would be to use a single-anode rectifier in series with each anode of a polyphase rectifier. Then before a short circuit could occur, an arc-back would have to occur in both rectifiers at the same instant or within the same period of about 5 microsec. The probability of this occurring is extremely small. In practice, this method is seldom used, for actually the probability of arc-back can be reduced sufficiently by the use of shields, grids, and cooling coils to satisfy most power applications, especially in the 600-volt range.

11-12. Thermionic Gas Diode: The Phanotron.—The hot-cathode gas-filled diode, or phanotron, usually has an oxide-coated cathode (heat-shielded¹ in the larger tubes) with an anode

¹ See Sec. 6-4.

at a distance of a few millimeters to several centimeters. The tube may be filled with one of the inert gases such as argon to a pressure of 0.1 to 0.5 mm. Hg, or it may contain a few drops of mercury so that the pressure corresponds to the vapor pressure of mercury at the condensation temperature in the tube, say 0.001 mm. at 18°C. to 0.1 mm. at 82°C. The discharge through a tube such as this has been classified as an arc (Sec. 11-2) because it has the characteristic low cathode fall of an arc. The total voltage drop across the tube is also low, about 10 volts in mercury tubes, and remains nearly constant with current variations within the operating range (Fig. 11-39). This feature contributes to the high efficiency and low tube heating of the

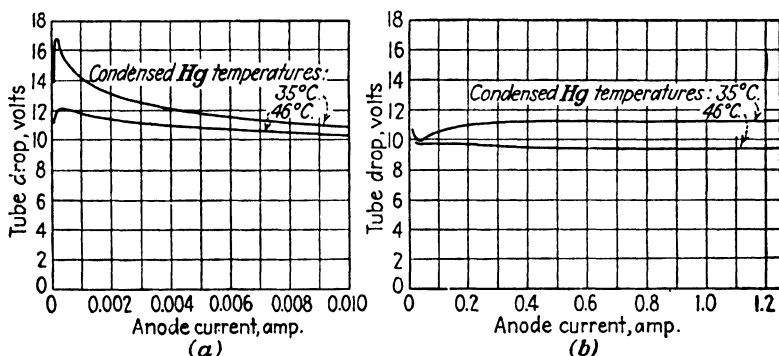


FIG. 11-39.—Tube drop across a mercury-vapor phanotron. (a) Low current range, (b) higher currents (FG-19A).

phanotron compared with the high-vacuum diode in rectifier applications.

A description of the operation of the tube can conveniently be divided into two parts. A brief account of the conditions in the tube as the current is increased from zero will be followed by a discussion of the conditions when the arc has been fully established.

As the voltage across the tube is increased from zero up to the first critical potential of the gas (4.6 volts for mercury), the current consists entirely of electrons emitted by the hot cathode. If the gas pressure is below about 0.1 mm. Hg the electron flow is not appreciably impeded by collisions with gas atoms, and as a consequence the current is limited by the electron space charge. The current is small on account of the relatively large cathode-

to-anode distance. When the tube voltage exceeds the first critical potential and approaches the ionization potential (10.4 volts for mercury) and some of the atoms are receiving energy in exciting collisions, ionization can occur by successive impacts. In this way a few positive ions can appear. These will have a profound influence on the current, for they will neutralize a portion of the negative space charge during their slow journey to the cathode. When the voltage exceeds the ionization potential of the gas, the rate of positive-ion formation rises rapidly and the

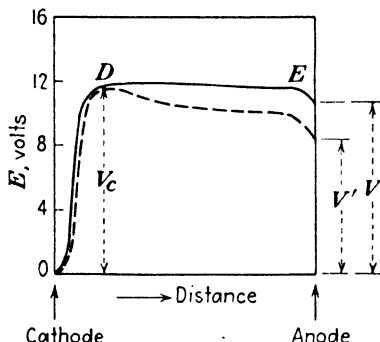


FIG. 11-40.—Potential distribution in a phenotron. Solid curve at a low pressure; dotted curve at a higher pressure.

space-charge neutralization increases correspondingly. The current will thus tend to increase without an increase in voltage; and, in general, the current must be stabilized by resistance in the anode circuit. In this condition the glowing region of the discharge resembles the positive column of a glow discharge; in fact, the interelectrode space, with the exception of thin layers or sheaths near the electrodes and walls, is

filled with a plasma of the type discussed in Sec. 10-6.

The potential distribution curves between the cathode and anode are of the type shown in Fig. 11-40. We observe that the greatest drop in voltage occurs near the cathode, whereas in the plasma, region *DE*, the change in voltage is small. The conditions at the anode are usually such that the anode fall is negative. As a result, the cathode fall of potential V_c can be higher than the anode-to-cathode voltage V . At the higher pressures a nonuniformity along the plasma contributes to this effect.

The conditions near the cathode are all-important in the determination of the tube voltage and in the maintenance of the discharge. The electrons from the hot cathode are being accelerated through the cathode fall V_c , so that they enter the plasma with high velocity. Here they lose energy by ionizing and exciting impacts, so that the necessary ionization is developed at the cathode end of the plasma. The positive ions are accelerated toward the cathode; and although some of them lose energy

in collisions with neutral atoms, an appreciable proportion bombard the cathode with an energy of about V_c e.v. In a theoretical treatment, Langmuir¹ finds that immediately adjacent to the hot cathode is a region where the net space charge is negative, similar to that in a high-vacuum tube, and that a region of net positive space charge separates this negative space-charge region from the plasma. This so-called "double" sheath leads to a potential distribution such as is illustrated in Fig. 11-41 for parallel-plane electrodes. The curve has a horizontal slope at the cathode because of the assumption of a very large supply of electrons of zero initial velocity from the cathode. The slope is also zero at the plasma because the positive-ion current is limited by the positive space charge in this region. There is an abundant supply of positive ions in the plasma, and just enough move toward the cathode to keep the electric intensity practically zero at the plasma boundary.

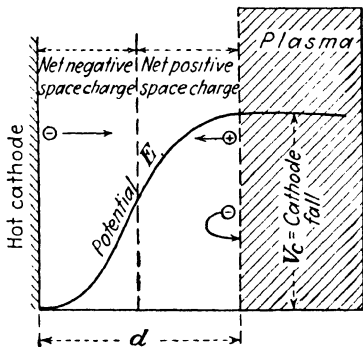


FIG. 11-41.—Potential distribution in the double sheath at the cathode.

Under these conditions and under the further assumption of a negligible effect of collisions, the electron current density is given by the expression

$$J_e = 1.86 \frac{4\epsilon}{9} \sqrt{\frac{2q_e}{m_e}} \frac{V_c^{3/2}}{d^2} \quad (11-39)$$

and the positive-ion current density is given by

$$J_p = 1.86 \frac{4\epsilon}{9} \sqrt{\frac{2q_p}{m_p}} \frac{V_c^{3/2}}{d^2} \quad (11-40)$$

where V_c is the cathode drop and d is the sheath thickness. These expressions differ only slightly from Eq. (6-35) for space-charge-limited current due to charges of one sign. The electron current is greater by a factor of 1.86 over what it would be in a high vacuum with the voltage V_c and the distance d . This increase is the result of the partial neutralization of the space

¹ LANGMUIR, I., *Phys. Rev.*, **33**, 954, 1929.

charge by the positive ions. The equations above make possible the calculation of the ratio of the electron current to the positive-ion current at the cathode, with the result that $J_e/J_p = \sqrt{m_p/m_e}$. For argon the ratio is 269; for mercury, it is 606. Thus, the current is carried almost entirely by electrons.

To give the order of magnitude of the space-charge-sheath thickness, suppose that d be calculated by Eq. (11-39). Let the current density J_e be 1,000 amp./sq. m., and let V_e be 10 volts, a reasonable value in mercury vapor. Then, d proves to be 0.37 mm. If V_e remains constant, the value of d will decrease to 0.12 mm. when the current density increases to 9,000 amp./sq. m.

It was not until experiments disclosed the conditions under which hot cathodes could be operated in a gaseous discharge without excessive cathode sputtering that the phanotron became a practical device. These experiments¹ showed that there is a critical tube voltage drop, called the *disintegration voltage*, above which cathode sputtering becomes serious. This voltage is rather definite for thoriated-tungsten cathodes, and its value depends upon the gas. For example, Hull gives the following values: 22 volts for mercury, 25 volts for argon, and 27 volts for neon. With oxide-coated cathodes the disintegration voltage is not so definite, but a value of 25 volts is representative. Ordinarily, then, the tube drop must be kept small.

The arc voltage drop is extremely important not only as a measure of the disintegration of the cathode but as a factor in the efficiency of operation as well. Under normal operating conditions, when the available electron emission of the cathode exceeds the current through the tube, the tube drop is mainly due to a cathode fall approximately equal to the ionization potential of the gas. Under abnormal conditions, as under short circuit or with a partly deactivated cathode, the zero-field electron emission of the cathode may be less than the current through the tube. Then the cathode fall will rise until there is an accelerating field for electrons at the cathode, and the current will increase by the Schottky effect to meet the demand. If the corresponding tube drop rises above the disintegration voltage, destructive positive-ion bombardment of the cathode will set in and the cathode coating will disappear.

¹ HULL, A. W., *Trans. A.I.E.E.*, **47**, 753, 1928. KOLLER, L. R., *Physics*, **7**, 225, 1936.

Another factor affecting the arc drop is the pressure of the gas. In tubes filled with an inert gas the pressure changes only slightly with the temperature; furthermore, the mean free path of the ions remains constant, so that little change in arc drop is observed. However, the cleanup of the gas will lower the pressure gradually during the life of the tube, and as a consequence the arc drop will increase. With mercury-filled tubes the large change of vapor pressure with temperature causes correspondingly large changes in the arc drop. This variation is illustrated in Fig. 11-42. The abscissas represent the so-called "condensed mercury

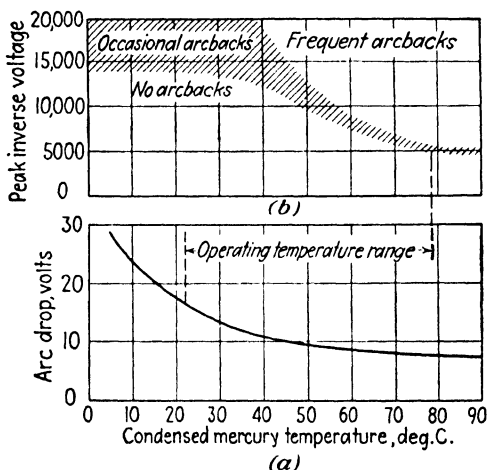


FIG. 11-42.—(a) Arc drop and (b) arc-back characteristics for a mercury-vapor phanotron.

temperature" as observed on the outside of the tube at its coldest point. It is clear that the tube cannot be operated below a certain limiting low temperature on account of the rise in the arc drop.

When used in rectifier circuits the phanotron is subject to backfire, or arc-back, similar to that in mercury-arc rectifiers (Sec. 11-11). At the higher gas pressures the arc-back is essentially a sparking phenomenon similar to that discussed in Chap. IX, with conditions corresponding to points to the left of the minimum sparking potential (Fig. 9-4a). Consequently, an increase in the pressure reduces the voltage at which arc-backs become numerous, as shown in Fig. 11-42. This leads to the limitation on the maximum temperature of operation of the

mercury tubes for a given voltage rating. At the lower gas pressures the random nature of the arc-back is more prominently displayed. The conditions in the gas tube are different from those in a tank rectifier in that there is no arc present during the nonconducting period, so that there is no ionization in the tube except immediately after the current drops to zero. About 0.001 sec. after the current zero the deionization is complete. In circuits in which the inverse voltage builds up rapidly an arc-back may occur before deionization is complete under conditions similar to those in the tank rectifier, as discussed in the preceding section.

Since any change in the anode surface favorable to the formation of a cathode spot is likely to lead to an arc-back, sputtering of the cathode with the accompanying deposition of barium and oxide particles on the anode will promote arc-back by providing places where thermionic emission can occur. Consequently, the temperature of the anode must be kept low. As in the mercury-arc rectifier an increase in the arc current causes an increase in the arc-back rate. Likewise, a low cathode temperature is conducive to arc-back.

To summarize, we may emphasize that the phanotron is characterized by a low and nearly constant tube drop during the conducting period in normal operation. Its filamentary or indirectly heated cathode can be made very efficient by the use of the heat-shielding principle, which entails, however, a cathode heating time of 30 sec. to several minutes or more. The cathode will operate satisfactorily as long as the arc drop does not rise above the disintegration voltage for the particular gas and cathode material. The arc drop will rise whenever the tube current exceeds the total electron emission of the cathode, so that the safe peak currents through the tube are limited accordingly. The factor that limits the operating voltage is the tendency to arc back when the anode is negative. In mercury tubes, both the arc drop and the probability for arc-back depend on the condensed mercury temperature, limits being thus set on the maximum and minimum operating temperatures. The heating of the anode and cathode must be limited to reasonable values, and consequently the average current through the tube is also limited to a definite rated value, depending on the design of the tube.

11-13. Thermionic Gas Triode: The Thyatron.—A gas-filled hot-cathode tube having one or two control electrodes is called a *thyatron*. The control electrodes are called *grids*, although it is only in the smallest thyratrons that the grid structure is similar to that in a high-vacuum triode. The electrode structures of two typical thyratrons are shown in Fig. 11-43a and 11-43b. The function of the grid in the control of the anode current is quite different in the thyatron than in the high-vacuum tube. The grid function is solely to *prevent the initiation of an arc* from cathode to anode until the grid and anode potentials have been

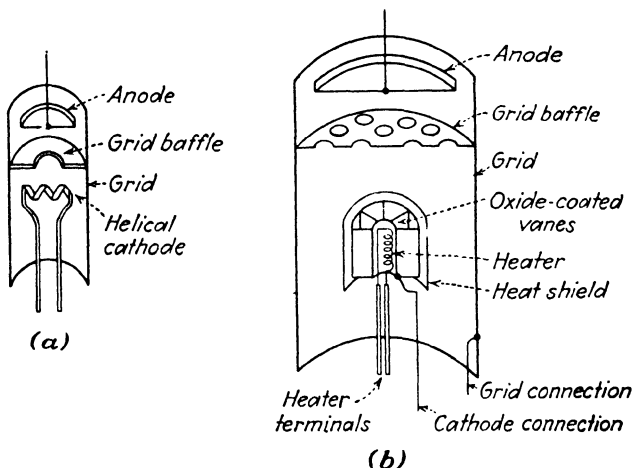


FIG. 11-43.—Electrode structures of thyratrons.

adjusted to cause breakdown. Once the arc is started, the grid loses its control function, and the anode voltage must be removed or reversed to permit the tube to deionize before the grid regains control.¹ In most thyatron applications, this occurs automatically; for the anode is fed from a source of alternating voltage, and the anode periodically goes negative with respect to the cathode.

The anode voltage required to cause breakdown depends on the grid voltage as shown in Fig. 11-44 for a small inert gas-filled thyatron. This curve is called the *control characteristic* or *breakdown characteristic* of the tube. Once the discharge starts, the tube behaves exactly like a phanotron as far as its

¹ Exceptions occur in small thyratrons where a large negative voltage on the grid may stop the discharge.

anode current is concerned. The arc drop is nearly independent of the current but is slightly higher than in an equivalent gas diode, because of deionization on the grid.

To consider the breakdown process in somewhat more detail, let us suppose that a grid voltage of a few volts negative is applied and that the anode voltage is gradually increased. At first, the retarding field of the negative grid will prevent the electrons emitted from the cathode from passing through the holes in the grid baffle. As the anode potential is raised, however, the field in the neighborhood of the holes in the grid changes,

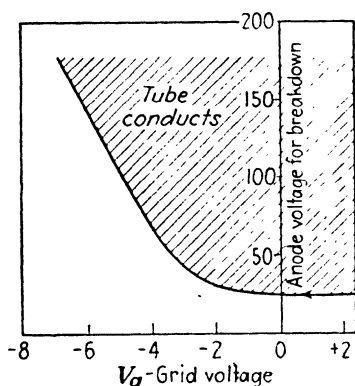


FIG. 11-44.—Control characteristic for an argon-filled thyatron.

so that some of the electrons in the cathode-grid space are accelerated through the holes toward the anode. Near the cathode the field still tends to retard the electrons, so that only the electrons emitted with a high velocity in the direction of the anode pass through the holes in the grid. These electrons are accelerated toward the anode and gain enough energy so that, when they collide with gas atoms, impact ionization takes place in the grid-anode space. The resulting

positive ions in the grid-anode space act to increase the field strength near the grid holes and to draw more electrons toward the anode. These processes are cumulative, and after the current reaches a critical value an arc is quickly established. Had the grid been more negative, a higher anode voltage would have been needed to draw enough electrons past the grid to initiate breakdown.

Once the arc is established, the negative grid behaves much like a probe in an ionized gas as discussed in Sec. 10-6. The grid is covered with a sheath having a net positive space charge which effectively neutralizes the negative potential on the grid, and the influence of the grid surface charges does not extend beyond the sheath boundary. In the usual thyatron, this sheath is so thin that it does not impede the flow of current in the main discharge, so that changes in grid voltage have no effect

on the anode current. In small thyratrons, however, the grid sheaths may become thick enough with large negative grid voltages to tend to overlap and thus shut off the anode current. This tendency is opposed by the increasing current density in the plasma remaining between the grid sheaths, for the accompanying increase in positive-ion current density entering the sheath tends to decrease the sheath thickness at this point. On account of these counteracting effects, it is only in small thyratrons carrying small anode currents that a negative grid can cut off the arc current. In most thyratrons the grid acts solely

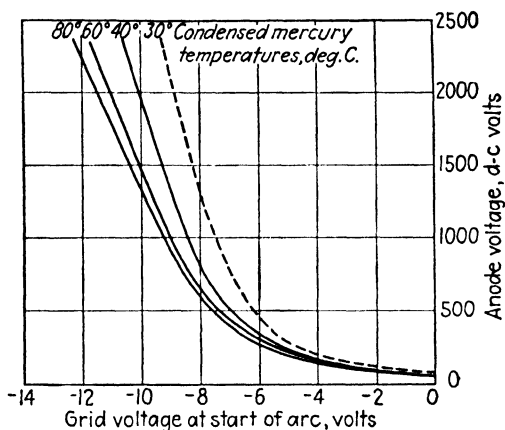


FIG. 11-45.—Control characteristics for a mercury-vapor thyatron at various condensed mercury temperatures (FG-17).

to initiate the cathode-anode arc, and after the arc is established it cannot control the anode current.

The foregoing discussion of thyatron breakdown indicates that the pressure of the gas will have an effect on the control characteristic, for the efficiency of the ionization processes depends on the gas pressure. The mean free path is about the same as the cathode-to-anode distance in the usual tube; thus, even though an electron has the requisite energy, its chance of producing ionization depends mainly on its chance of having an impact. Therefore an increase in pressure will increase the probability of collisions, and the ionization rate; consequently, the critical current will be reached at a lower anode voltage. Thus, the breakdown voltage will depend on the vapor pressure as shown in Fig. 11-45 for a mercury-vapor thyatron.

The so-called *positive-grid thyatron* has been designed so that positive voltages are needed on the grid to start conduction. In these tubes the grid is made more effective as a shield, for example, by arranging two or three baffles, of the type shown in Fig. 11-43*b*, one above the other. Then the anode voltage produces little or no accelerating field between the cathode and the first grid baffle. As a result, the grid potential must be a few volts positive to accelerate enough electrons through the grid baffles to start the ionization leading to breakdown. Furthermore, the grid voltage required for breakdown is nearly independent of the anode voltage, so that the control characteristic approximates a vertical line.

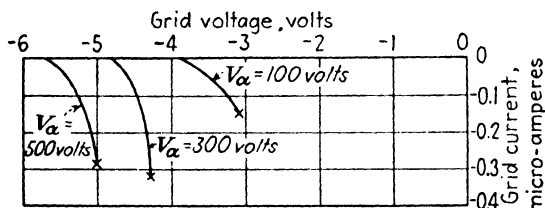


FIG. 11-46.—Grid currents preceding breakdown in an argon-filled thyatron. V_a indicates anode voltage. Breakdown occurs at x .

A knowledge of the grid-current characteristics of thyatrons is of importance, especially when the grid is controlled by high-resistance circuits as, for example, circuits involving phototubes. The grid currents can be discussed under two heads, (1) those preceding the tube breakdown and (2) those flowing after an arc is established. In negative grid thyatrons the first class of grid current arises from (a) leakage over grid insulation, (b) thermionic emission of electrons from the grid, and (c) positive ions drawn from the pre-breakdown discharge in the tube. Thermionic emission is small if the grid does not heat excessively and if it is kept free of the oxide from the cathode. The grid current caused by positive ions is similar to that in a gassy vacuum triode (Sec. 6-9). The grid currents up to breakdown for an argon-filled thyatron are shown in Fig. 11-46. The grid current following breakdown is essentially the same as the current to a probe in an ionized gas (Sec. 10-6). The grid-current curves for a mercury-vapor thyatron with the arc present are shown in Fig. 11-47. It should be noted that the grid current becomes very large when the grid is positive, for the grid can collect

electrons from the arc. It is usually necessary to provide a resistance in series with the grid to limit this current in circuits where the grid potential may become positive.

In the so-called *double-grid*, or *shield-grid*, thyatron the pre-breakdown grid current has been materially reduced by the

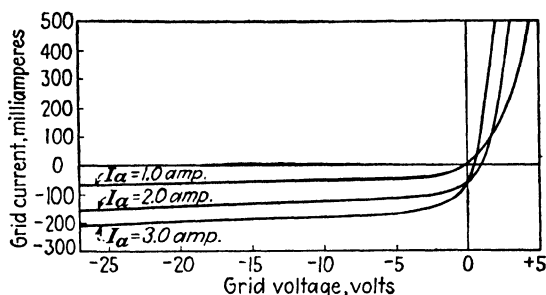


FIG. 11-47.—Grid currents after breakdown in a mercury-vapor thyatron (FG-57). I_a indicates direct current in the anode circuit. (Reprinted by permission from E. D. McArthur, "Electronics and Electron Tubes," p. 129, John Wiley & Sons, Inc., New York, 1936.)

addition of a second grid. One form of its construction is shown in Fig. 11-48. When the shield grid is connected to the cathode, the control characteristic is very similar to that in Fig. 11-45 for the negative-grid thyatron. However, the grid currents preceding breakdown are reduced about a hundredfold compared with a similar single-grid thyatron; thus, the grid can readily be controlled through a high-impedance circuit. An additional advantage of the shield-grid thyatron is that the control grid has low values of anode-to-grid and grid-to-cathode capacitance.

Although the time required to establish conduction, called the *ionization time*, is only a few microseconds, the *deionization time* may be 10 to 1,000 microsec. The deionization time is defined as the time required for the grid to regain control after removal of the anode voltage. This time is important in applications where the tube must operate at a high frequency or where the anode voltage goes positive soon after a conduction period, as in inverter circuits. For obvious reasons, the deionization time is

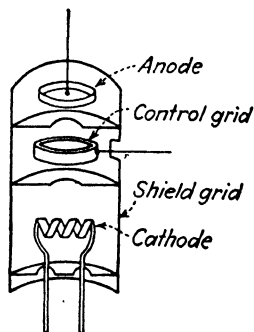


FIG. 11-48.—Structure of a double-grid thyatron.

smaller in small tubes having short interelectrode spacings. Other factors that decrease the deionization time are the following: smaller load current, reduced gas pressure, and a more negative grid potential.

The thyatron tube has ratings similar to those of the phanotron for maximum anode currents, maximum inverse anode voltage, etc. In addition, it has a rated limitation on grid current, on grid voltages, and on the maximum positive voltage on the anode, called the maximum *forward* voltage. The value of the latter is the maximum value consistent with reliable control of the discharge by the grid.

11-14. Thyatron Applications.—A highly diversified list of thyatron applications has been developed, and more are constantly appearing. A classification of the main applications as to function is given in Table 11-3.

TABLE 11-3.—THYRATRON APPLICATIONS*

Function	Examples of Application
Power amplification and relaying	Operation of circuits from phototubes Timer circuits Remote control
High-speed circuit closure and interruption	Control of high-speed spot and seam resistance welding
Circuit impedance control.....	Theater stage and auditorium lighting Voltage control of a.c. and d.c. generators
Current commutation.....	A.c. to d.c. rectification D.c. to a.c. inversion Frequency changing Thyatron commutator variable speed a.c. motors

* WHITELEY, A. L., *J.I.E.E.*, **76**, 516, 1936.

In all these applications the thyatron functions as a tube that can pass a unidirectional current with a relatively low voltage drop after the grid allows the discharge to start. It must suffice to describe a few applications involving generally useful basic circuits.

A description of the single-phase thyatron rectifier with control of the rectified current by means of the amplitude of the d.c. grid voltage will serve to illustrate a general method of analysis. The circuit diagram is shown in Fig. 11-49, and let us assume that the control characteristic of the thyatron is that shown in the left-hand portion of Fig. 11-50. For the moment let us assume that the voltage wave of the a.c. source, $E_m \sin \omega t$

in Fig. 11-50, is impressed directly on the anode of the tube. This will actually be the case as long as there is no current in the anode circuit, *i.e.*, before the tube breaks down or "fires." To analyze the action of the grid, it is convenient to construct a curve showing the instantaneous voltage of the grid that would just allow breakdown at every instant of time during the positive half cycle of anode voltage. This can be done by taking grid-voltage values for different anode voltages from the control characteristic and plotting them in the time diagram or by using the graphical construction indicated in Fig. 11-50. This construction consists in transferring the critical grid voltage for a particular anode voltage from the control characteristic to the time diagram at the instant at which the anode voltage has this particular value.

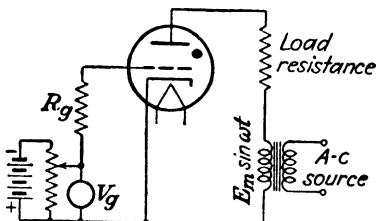


FIG. 11-49.—Thyatron rectifier controlled by the magnitude of the grid bias voltage.

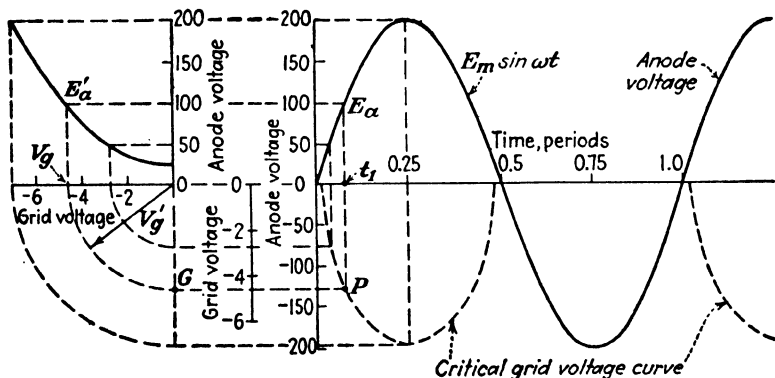


FIG. 11-50.—Diagram showing a construction for the critical grid-voltage curve. For example, to locate the critical grid voltage at instant t_1 project upward to the anode voltage E_a . Then project horizontally to point E_a' on the control characteristic. This locates the critical grid voltage $V_{g'}$. To transfer $V_{g'}$ to the time diagram swing an arc of radius $V_{g'}$, indicated by $V_{g'}$ in the diagram, and project horizontally from G to point P , the required point.

To analyze the actual operation of the rectifier, we make use of the critical-grid-voltage curve as shown in Fig. 11-51. We assume that the resistance R_g is small enough so that the pre-breakdown grid currents will not alter the grid potential. When

the constant negative grid potential V_g exceeds the critical grid voltage (at intersection x , Fig. 11-51), the tube will fire. The anode voltage meanwhile rises along the sine curve of the source voltage until the instant of firing t_1 and then drops to the characteristic low arc drop. At the instant of firing the load current jumps suddenly to a value determined by the instantaneous source voltage, the load resistance, and the arc drop. Load current continues to flow for the rest of the half cycle to the instant t_2 . During the ensuing half cycle the tube deionizes, and firing recurs at instant t_3 .

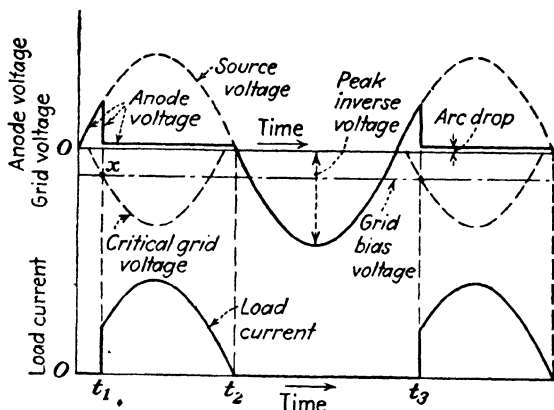


FIG. 11-51.—Anode voltage and load-current wave forms in a thyatron rectifier controlled by negative grid bias.

Upon increasing the grid bias V_g , the instant of firing shifts to later instants of time, until the tube ceases firing when the grid bias exceeds the negative maximum of the critical-grid-voltage curve. Thus, control of the instant of firing is had over the range of the first quarter period of the source voltage wave, with a corresponding range of control of approximately 2 to 1 over the average value of the anode current. With mercury thyatrons the critical-grid-voltage curve will change as the tube temperature changes, with corresponding changes in the instant of firing. For this reason the control of mercury thyatrons by the amplitude of the grid bias is most suitable for "on-off" control, where the grid bias changes are made large enough to ensure stopping or starting the discharge.

An alternative and more flexible method of controlling the output of the thyatron rectifier is that of feeding the grid with

an alternating voltage whose phase position relative to that of the anode voltage can be varied. Several methods of shifting the phase of the grid voltage are available.¹ In the method shown in Fig. 11-52 a variation in the resistance R changes the phase of the voltage from K to g , but the magnitude of this voltage remains constant. To study the effect of this change in the phase of the alternating grid voltage, consider the waves of grid voltage in relation to the critical-grid-voltage curve for the three phase positions shown in Fig. 11-53*a*, *b*, and *c*. At *a* the grid voltage is 180 deg. out of phase with the anode voltage (R very large), and the thyatron is prevented from firing. At *b* the

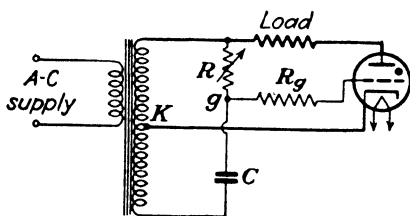


FIG. 11-52.—Circuit diagram of a phase-controlled thyatron rectifier.

grid-voltage wave has been advanced 60 deg., and firing occurs at instant t_1 where the actual grid voltage exceeds the critical grid voltage. Thus the tube begins to conduct for a small part of the cycle, and the average current is small. By advancing the phase of the grid voltage still more, as shown in Fig. 11-53*c*, firing takes place earlier in the cycle and the average current is increased. Thus a continuous control of the output is had by changing the resistance R from a large value down to zero. The function of R_g is to prevent excessive grid currents during the conducting period, but its value should be kept as small as possible to avoid an excessive voltage drop over R_g , caused by pre-breakdown grid currents.

The control of an alternating current by two thyratrons is exemplified by the welding control circuit shown in Fig. 11-54. The schematic diagram in Fig. 11-54*a* shows the general plan. The current to the primary of the welding transformer is controlled by a series transformer with the associated thyratrons. When these thyratrons are rendered nonconducting by a suitable grid voltage, the series transformer offers a high impedance to

¹ General reference 11-9, pp. 445-452.

current flowing to the welding transformer. As a result, only a small current flows. However, when the tubes are both rendered conducting, the current through the series-transformer secondary is limited only by the low tube voltage drop. The series transformer acts as if its secondary were short-circuited,

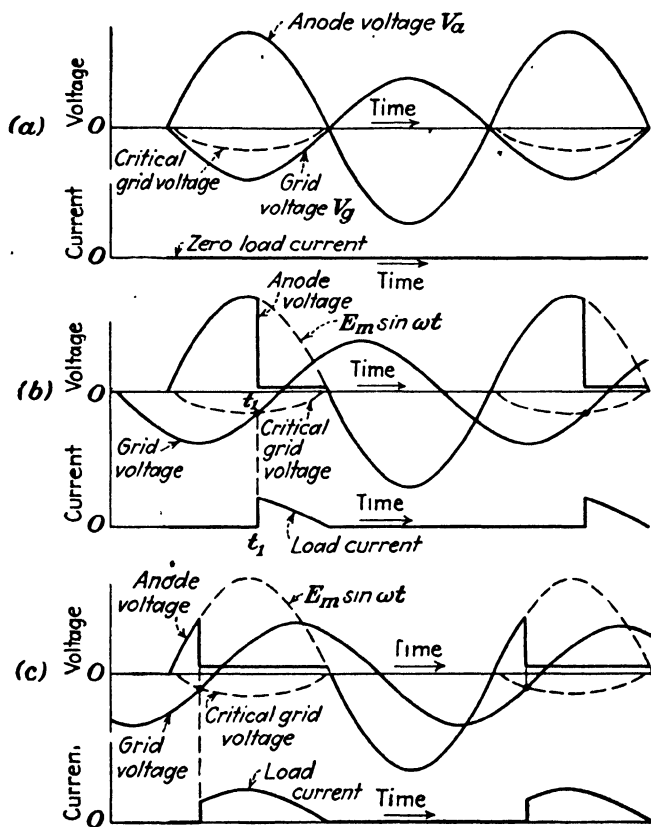


FIG. 11-53.—Anode voltage, grid voltage, and load-current waves in a phase-controlled thyatron rectifier circuit. (a) V_g lags V_a by 180° (b) V_g lags V_a by 120° . (c) V_g lags V_a by 60° .

and consequently the welding-transformer primary receives nearly the full voltage of the line. The tubes handle currents much smaller than the line currents but must withstand voltages higher than line voltage. It is more economical to use the low-current high-voltage tubes in the circuit shown than to attempt to use high-current tubes directly in the power circuit.

The details of a grid control circuit employed for providing "on-off" control for seam welding by means of overlapping spot welds are shown in Fig. 11-54b. In a typical welding cycle a current of the order of 1,000 amp. flows to the primary of the welding transformer for a period of 0.033 sec., and this is followed by an "off" period of 0.033 sec. With the control switch in

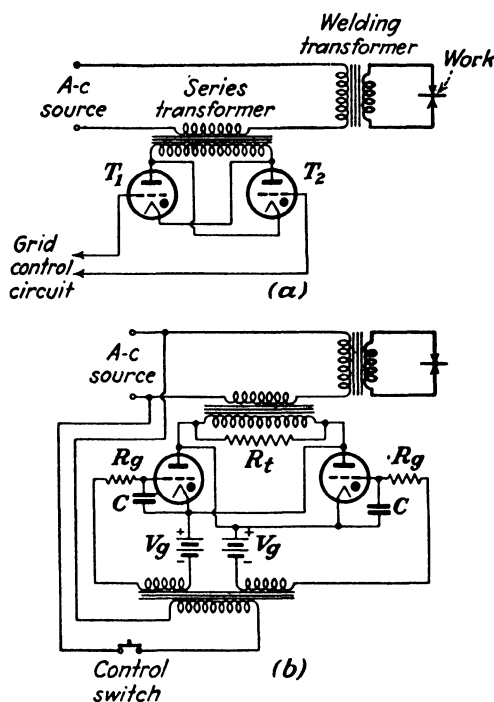


FIG. 11-54.—Thyatron welding control circuits. (a) Schematic diagram. (b) Details of grid control circuit. (Palmer, H. L., *Gen. Elec. Rev.*, 40, 230, 1937.)

Fig. 11-54b open, the bias voltages V_g prevent conduction through the thyratrons. When the control switch is closed, alternating voltages are impressed on the grids of both tubes, each in phase with the anode voltage of that tube. The magnitude of these voltages is adjusted so that conduction starts early in the cycle. Condensers C serve to prevent surges from firing the tubes. In practice the bias voltages V_g are supplied by small rectifiers. The *Thyrille* resistor R_t provides protection against surges that originate from switching transients or other causes.

The control switch in Fig. 11-54b is replaced by a thyatron timing circuit when accurately reproducible welds are desired. These timing circuits take various forms, but usually the timing is accomplished by feeding the grid of a thyatron from a circuit in which a condenser discharges through a resistance. The time to fire the thyatron can be adjusted to predetermined values by a suitable choice of the constants of the condenser discharge circuit.

Another method of controlling an alternating current is by the use of a series saturable reactor and a full-wave controllable thyatron rectifier. The elementary diagram of this circuit

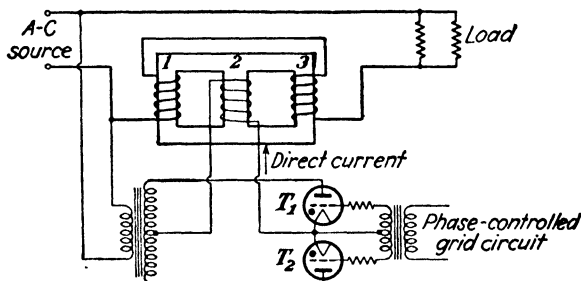


FIG. 11-55.—Circuit for controlling an alternating current by means of thyatrons and a saturable reactor.

is shown in Fig. 11-55. The load is fed through the coils 1 and 3 of the saturable reactor, and the thyatron tubes T_1 and T_2 provide a controllable direct current through the coil 2. The coils are arranged in such a way that no alternating flux traverses the central core. When the direct current through coil 2 is small, the reactance of coils 1 and 3 is high, and thus the current to the load is small. However, when the direct current is large, the core reaches magnetic saturation so that the reactance of coils 1 and 3 falls. Consequently, a large load current can be controlled with stepless variation by means of a few watts of power expended in the grid control circuit. The maximum a.c. power in the load can be 100 times the d.c. power fed to the saturable reactor, and the total power losses in the saturable reactor may be kept down to 2 to 3 per cent of the maximum power in the load. This method of control has been applied to theater-stage and auditorium lighting.¹

¹ "Electrical Engineers' Handbook," 3d ed., Vol. V, Sec. 16, pp. 22-26, John Wiley & Sons, Inc., New York, 1936. GULLIKSEN, F. H. and E. H.

The function of current commutation, or the shifting of a current from one circuit to another, is performed by the two thyratrons in the full-wave rectifier in Fig. 11-55. Another example of current commutation by thyratrons will be discussed in Sec. 11-16 in connection with d.c. to a.c. inversion.

11-15. The Ignitron.—The ignitron was developed in an effort to produce an improved pool type of mercury-arc rectifier. The

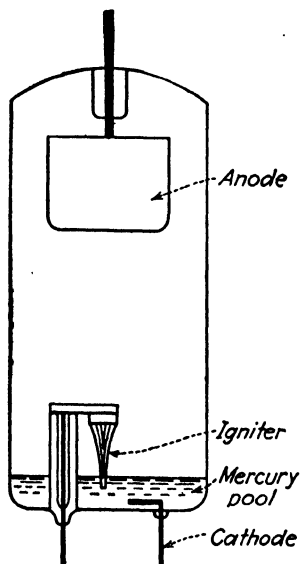


FIG. 11-56a.—Diagram of a glass-bulb ignitron.

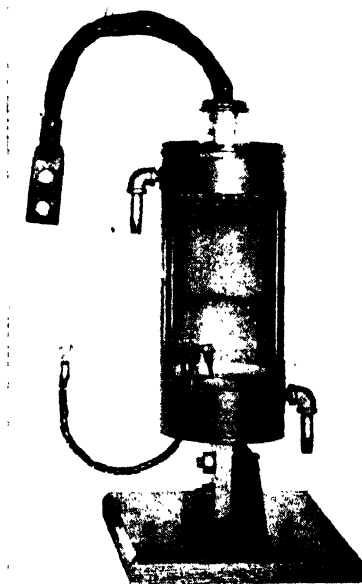


FIG. 11-56b.—Cutaway view of a water-cooled metal ignitron, Type FG-235A. (Courtesy of the General Electric Company.)

basis for the development was the fact that the presence of the arc in a multianode tank rectifier promoted backfires and if each rectifying path could be isolated from the others the backfire rate would decrease. However, a single tube containing a mercury-pool cathode and a conventional anode could rectify alternating current only if some means could be found to initiate a cathode spot on the mercury pool at the beginning of each

conducting period. In 1933, Slepian and Ludwig¹ announced a new method for initiating a cathode spot. The method is to send a current into the mercury pool through a so-called "igniter" rod, which consists of a small conical rod of boron carbide or silicon carbide (carborundum) with its pointed end dipping into the mercury. The ignitron, as shown in Fig. 11-56, consists of a highly evacuated tube containing a mercury-pool cathode with an igniter rod dipping into it and having another insulated electrode which serves as the anode. In the small sizes the tubes are of the air-cooled glass type; larger tubes are sealed-off metal tubes having water cooling (Fig. 11-56b); the largest tubes are demountable and must be pumped continuously.

Slepian and Ludwig² have explained the action of the igniter in terms of the concentration of current at the point where the igniter rod makes contact with the mercury surface. As the current leaves the rod and enters the more highly conducting mercury, it tends to crowd toward the contact line at the mercury surface. This concentration of current produces a voltage drop across a small gap near the rod-mercury junction sufficient to lead to the formation of a cathode spot on the mercury surface. The conditions at the junction are similar to those present when an arc is formed at separating contacts. In practice, an igniter current of 10 to 20 amp. at a voltage of about 100 volts is sufficient to initiate a cathode spot. With the initiation of the cathode spot an arc forms almost immediately between the anode and the cathode pool.

The experiments of Dow and Powers³ show that a finite, though variable, time elapses after the igniter current is established before the arc is complete from anode to cathode. For example, in one series of tests, 65 per cent of the starting times were less than 10 microsec., and 94 per cent were less than 25 microsec. The median time was about 7 microsec. The starting time is reduced with higher igniter current.

For the ignitron to function as a rectifier the arc must be ignited at the beginning of each conducting period by a pulse of current given to the igniter. At the end of the conducting period the arc goes out, and after a short deionization time the

¹ *Trans. A.I.E.E.*, **52**, 693, 1933.

² *Loc. cit.*; also, see J. M. Cage, *Gen. Elec. Rev.*, **38**, 464, 1935.

³ *Trans. A.I.E.E.*, **54**, 942, 1935.

tube acts as an insulator. One scheme for exciting the igniter consists in connecting the igniter through a small phanotron tube to the anode of the ignitron, as shown in Fig. 11-57. When the ignitron anode potential goes positive, current flows through the phanotron to the igniter until the arc is struck. The anode-to-cathode voltage drop then decreases to about 12 to 15 volts, so that conduction through the auxiliary tube stops. Thus, the auxiliary rectifier carries current only during a small percentage of the positive half cycle until conduction is established in the ignitron.

If the impulse of current to the igniter is withheld until a later part of the positive half cycle, the firing of the ignitron

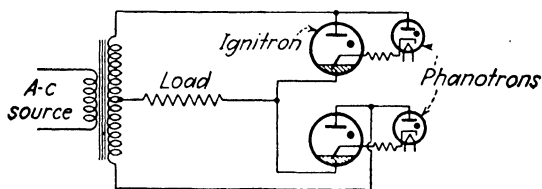


FIG. 11-57.—Full-wave ignitron rectifier circuit.

will be delayed correspondingly. Thus, the ignitron has all the possibilities of controlled rectification possessed by the thyatron, and by controlling the instant of igniter excitation it is possible to control the output of the ignitron rectifier in much the same fashion as in the phase-controlled thyatron rectifier. For example, in the rectifier in Fig. 11-57 the auxiliary rectifier tubes can be replaced by thyratrons. If the grids of these thyratrons are supplied from a phase-shifting circuit, it is easy to control the instant of firing of the thyratrons and hence of the ignitrons. By this method, load voltage control is obtained on large six-tube ignitron rectifiers.¹

The ignitron has the large momentary current capacity characteristic of the mercury-pool cathode, but its average current rating is limited by tube heating. Its high short-time current rating is an advantage when the tube is used in spot- and seam-welding control circuits,² an application in which it is supplanting thyatron control. The control circuit employed (Fig. 11-58) is similar to that in Fig. 11-54a except that no series

¹ COX, J. H., and G. F. JONES, *Trans. A.I.E.E.*, **58**, 621, 1939.

² PACKARD, D., and J. H. HUTCHINGS, *Elec. Eng.*, **56**, 37, 1937.

transformer is used. The impulses of welding current can be very accurately controlled as to duration by a thyatron timer circuit, incorporated into the igniter control circuits, and as to magnitude, if desired, by phase control of the firing impulses.

The ignitron combines the high instantaneous current capacity of the pool type of rectifier with a control feature similar to that of the thyatron. The backfire rate is smaller in the ignitron than in an equivalent pool type of rectifier because no arc plays in the tube while the anode is negative. The igniter control requires much higher instantaneous control power than for the grid control of a thyatron. Consequently, the ignitron is applied where the controlled power (load) is large, so that the power loss in the igniter circuits is relatively small. The ignitron has the advantage over the thyatron in that no cathode heating

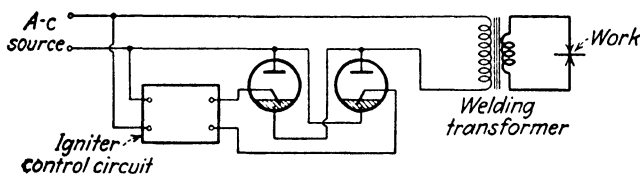


FIG. 11-58.—Schematic diagram of an ignitron welding control circuit.

power is needed and no delay for preliminary cathode heating is required. In the thyatron, the grid acts to *prevent the discharge*, whereas in the ignitron the discharge *cannot possibly start* until the igniter carries current. Therefore, the mechanism of the control of the two tubes is quite different, though they are applied in similar circuits.

11-16. Inverters.—A circuit for the production of alternating current from a source of direct current is called an *inverter*. The vacuum-tube oscillator accomplishes the same function, but at a relatively low efficiency. The use of thyatrons and ignitrons in inverter circuits opens possibilities of the conversion of large blocks of power at high efficiency. This possibility is attractive in view of the advantages of the d.c. system of power transmission in which inversion to alternating current is needed at the receiving end of the line to facilitate the distribution of the power over existing a.c. systems. Because of economic and technical limitations, inverters have as yet been commercially applied only in the small sizes.

In all inverter circuits the period of conduction through the gas tube must be terminated by forcing the anode potential¹ to zero or to negative values for a time long enough for the control device (grid or igniter) to regain control. To show how the anode current may be stopped in a thyatron supplied from a d.c. source, we shall describe the operation of the circuit in Fig. 11-59. The two-tube parallel type inverter will be treated as an extension of the principle involved in the operation of this circuit.

Let us assume that the thyatron has been ignited, say by momentarily closing switch S to bring the grid to cathode potential. Further assume that the bias voltage V_g is adequate to prevent breakdown when the anode voltage is raised to equal that of the d.c. source. With the tube conducting, the condenser C will charge up to the voltage across the load resistance with the polarity shown in the figure. If, now, switch S_1 is closed, the anode potential suddenly goes negative by an amount equal to the condenser voltage. However, the charge of the condenser decreases and eventually builds up with the opposite sign, so that the anode voltage rises, and becomes positive. If the tube deionizes sufficiently while its anode is negative, the grid will regain control, and the discharge will not reignite.

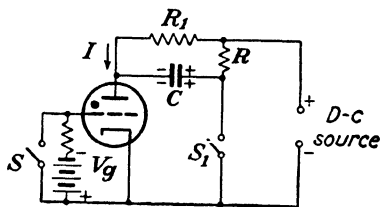


FIG. 11-59.—A circuit for stopping the anode current of a thyatron fed from a d.c. source.

In the two-tube parallel type inverter shown in Fig. 11-60 the discharge in one thyatron is stopped by the starting of the other thyatron. The operation of the circuit is similar to that of Fig. 11-59, except that the second thyatron replaces the switch S_1 . The condenser C in Fig. 11-60 makes possible the commutation of the current from one tube to the other. For example, assume that tube A is conducting and tube B is insulating. During this part of the cycle of operation, current flows from the d.c. source through the transformer winding from t to a , while a smaller current flows through the winding from t to b

¹ In the discussion, it will be assumed that all potentials are referred to the cathode of the tube.

to charge the condenser C . The anode of tube B is positive, but it does not fire until its grid voltage is raised by the positive loop of the a.c. component of grid voltage supplied by the external control source. When tube B does fire, the potential of the anode of tube A suddenly goes negative, and thus A is allowed to deionize. As the condenser C charges in the opposite direction, the anode potential of A rises, and thus tube A is ready to repeat the sequence just completed by tube B . The current is commutated back and forth between the two parallel paths through the two tubes, and consequently an alternating current is produced in the load.

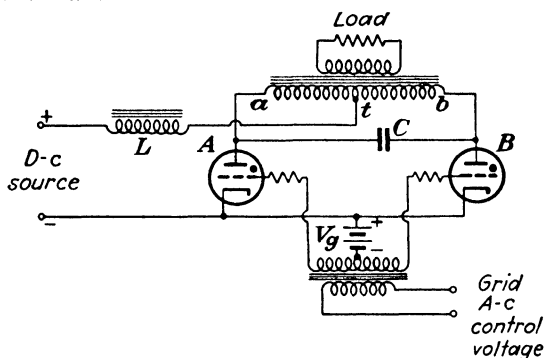


FIG. 11-60.—Externally controlled, parallel-type inverter circuit.

The waves of current in the d.c. circuit, anode voltage, anode current of tube A , and of load current are shown in Fig. 11-61 for a medium load where the inductance in the d.c. circuit is large. With light loads the load current wave is more peaked, and with heavy loads the wave is flat-topped. The time t_1 in the anode-voltage wave shows the time during which tube A must deionize. If tube A fails to deionize completely in this interval, it may begin to conduct again at the end of interval t_1 . If this happens, condenser C loses its charge, and the opposing magnetic effects of the two anode currents drawn through the windings $t-a$ and $t-b$ neutralize each other. As a result, the only opposing e.m.fs. in the series circuit of the d.c. source, the transformer, and one of the tubes are the small tube drop and the e.m.fs. of resistance of the transformer and the inductor L . The large current (essentially a short-circuit current) that would result would be destructive to the tubes, and so protection in the form of a quick-acting circuit breaker should be provided.

When the inductance L is made smaller, the current in the d.c. circuit may flow in pulses,¹ with intervening periods of zero current. In this condition, each tube conducts for less than half the time, and the tube is extinguished by a negative voltage arising in the transformer rather than by the mechanism previously discussed.

Wagner² has shown that with inductive loads the wave form of the load current is better than with resistance loads. He also discusses the time available for deionization in this case.

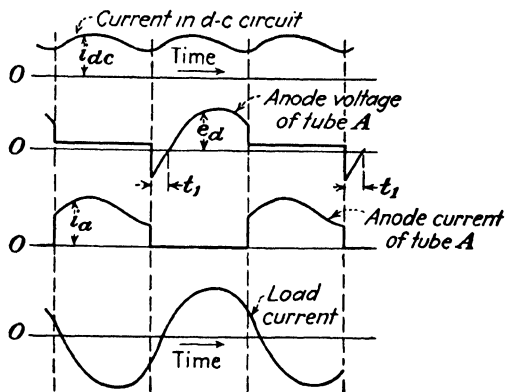


FIG. 11-61.—Wave forms of currents and voltages in the inverter of Fig. 11-60.

Part or all of the capacitance C (Fig. 11-60) can be placed across the secondary of the output transformer and still provide satisfactory commutation. Analysis of the circuit has shown that large values of commutating capacitance improve the output wave form and make the circuit more stable. However, a large capacitance makes the voltage regulation of the output poorer, so that in the choice of the condenser a compromise must be effected between the desired wave form and regulation characteristics.

Inverters proposed for high-power outputs usually have multiphase output circuits. One particularly interesting inverter³ converts constant-current d.c. power into constant-potential

¹ WAGNER, C. F., *Elec. Eng.*, **54**, 1228, 1935. TOMPKINS, F. N., *Trans. A.I.E.E.*, **51**, 707, 1932.

² *Elec. Eng.*, **55**, 970, 1936.

³ WILLIS, C. H., B. D. BEDFORD, and F. R. ELDER, *Elec. Eng.*, **54**, 102, 1935.

three-phase a.c. power. Another proposed use¹ for an inverter is in connection with an output rectifier to take d.c. power at one voltage and deliver d.c. power at a different voltage, *i.e.*, to provide a d.c. transformer circuit.

11-17. Gaseous-discharge Lamps.—The invention of the Geisler tube in 1850 marks the beginning of the scientific study of the light emitted from gaseous discharges. These studies have been concerned particularly with the spectral character of the light, *i.e.*, with the spectral lines and frequency bands in the radiation, and with the correlation between the observed spectra and the quantum theory of the atom. The quantum theory successfully accounts for the frequencies of the spectral lines of any atom in terms of the changes in the energies of its internal electrons.

In general, the radiation includes frequencies in the ultraviolet and infrared regions as well as in the visible range. For use as practical light sources, it is important to have as high a proportion of the energy as possible radiated in the visible range. Therefore, it is desirable to have the spectral lines and bands in the visible range as intense as possible and to adjust the discharge conditions to attain this end. Some knowledge of the way in which the strength of the spectral lines depends on the discharge conditions has been gained from a study of the atomic energy levels (internal electron energies) and of the energies of the electrons in the discharge for the simpler atoms,² but for other discharges the complexity of the factors operating compels recourse to experimental methods.

The conduction of the current and the generation of light are in a sense separate processes in the discharge; for the current is conducted as a result of the ionizing processes, whereas the light results from excitation of the atoms by electron impact. The conduction processes are, of course, necessary to provide the electrons in proper concentrations and with proper energies to produce the exciting impacts. Since both ionizing and exciting impacts abstract large amounts of energy from the discharge and only the excitation processes lead to useful light output, it can

¹ HERSKIND, C. C., *Elec. Eng.*, **56**, 1372, 1937.

² DUSHMAN, S., *J. Opt. Soc. Am.*, **27**, 1, 1937. FOUND, C. G., *Trans. Illum. Eng. Soc. (N.Y.)*, **33**, 161, 1938. See general reference 11-11.

be seen that the ionization should not be more intense than necessary for the optimum rate of excitation of the atoms.

As a preliminary to the discussion of typical modern gaseous lamps, we may classify them in several ways. (1) They may be classified as to the type of discharge as follows: (a) cold-cathode glow, (b) cold-cathode arc, and (c) hot-cathode (thermionic) arc. An example of the cold-cathode glow is the neon tube commonly used in advertising signs. An example under *b* is the Cooper-Hewitt mercury lamp with its mercury-pool cathode. Under *c*, we find most of the recently developed lamps, the sodium lamp, the high-pressure mercury lamp, and the fluorescent lamp. (2) A second classification is based on the source of the light, (a) positive column, (b) cathodic or negative glow, and (c) fluorescence excited by radiation from a positive column. Lamps of type *a* consist of long tubes of relatively small diameter; those of type *b* have a short spacing between the electrodes, compared with the bulb diameter. (3) Lamps may also be classified as to their operating pressure; thus we have low-pressure lamps (30 mm. Hg and below) and high-pressure lamps (say 300 mm. Hg and above).

a. Low-pressure Cold-cathode Glow Tube (Positive Column Type).—A description of the neon discharge tube used in advertising signs will serve to illustrate the principles involved in the use of cold-cathode glow tubes of the positive column type. These tubes are 1 to 20 m. in length and have diameters of 0.7 to 1.5 cm. An enlarged tubulation at each end houses the electrodes, which are usually hollow cylinders. Since the tube is fed with alternating current, the discharge disappears at the current zero and must be reignited each half cycle. The reignition voltage, which depends on the pressure of the gas and the length of the tube, will be of the order of 1,000 to 10,000 volts, so that the transformer must supply a voltage high enough to ensure breakdown of the gap. On account of the falling volt-ampere curve of the glow (Fig. 10-3), it is necessary to limit the current to stabilize the discharge. In these tubes the current is held to a value of 20 to 50 ma. by means of a high-leakage-reactance transformer, with the core and windings disposed as shown in Fig. 11-52.

The necessity of a relatively long tube in the interest of high luminous efficiency is clear when we consider the magnitude of

the cathode fall of potential in the glow. For example, if the cathode fall of potential is 200 volts, then the total tube voltage must be 1,000 volts if 80 per cent of the input power is to go to the positive column where the light is produced, and the tube must be long enough to provide about 800 volts drop along the positive column. The percentage of the power going to the positive column that appears as visible radiation depends on the losses, particularly radiation outside the visible range and energy losses to the walls (see Sec. 10-4), but in any case is not over a few per cent.

The optimum pressure from the viewpoint of light production is 1 or 2 mm. Hg, but tubes filled to this pressure have short

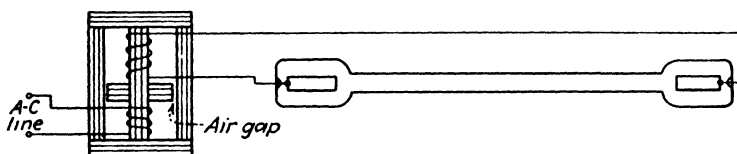


FIG. 11-62.—Cold-cathode glow tube fed by a high-leakage-reactance transformer.

lives on account of the "cleanup" of the gas caused by cathode sputtering. Hence, the tubes are filled to a pressure of 5 to 10 mm. to provide for adequate life. The breakdown voltage in these tubes is less than it would be for pure neon, because of the small percentage of argon usually mixed with the neon (see Sec. 9-4). As the tube nears the end of its life, the decrease in gas pressure and accompanying increase in the reignition voltage cause intermittent operation.

Some of the colors available in tubes of this type besides the red-orange color characteristic of neon are a pinkish white produced in helium, and a blue in a mixture of mercury vapor with argon and neon gas. Other colors are obtained by enclosing the discharge in colored glass tubing.¹

b. Low-pressure Hot-cathode Arc Lamps.—An example of this type of lamp is the sodium-vapor lamp, which is made in both the cathodic and the positive-column form. Figure 11-63 shows the positive-column lamp connected in a series circuit such as is used in highway lighting.

¹ For further information see "Neon Signs," S. C. Miller and D. G. Fink, McGraw-Hill Book Company, Inc., New York, 1935.

The vapor pressure of the metallic sodium in the tube is so low at ordinary temperatures that no discharge would be possible except for the fact that neon gas at a pressure of 3 mm. Hg is present. After the oxide-coated cathodes *C* (Fig. 11-63) are subjected to a preliminary 40-sec. heating, the relay opens and a neon discharge ensues. The molybdenum anodes *A* collect the current on one half cycle while the cathodes carry the current on the other half cycle. During the following 15-min. warming-up period the discharge gradually changes from the orange-red of neon to the characteristic yellow of the *D*-lines of sodium. The bulb operating temperature of about 235°C. is attained by insulating the bulb with a double-walled vacuum jacket.

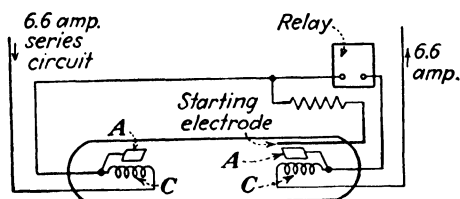


FIG. 11-63.—Constant-current sodium lamp circuit.

Under operating conditions the sodium vapor pressure is about 0.001 mm.; the neon atoms thus outnumber the sodium atoms in the ratio of 3,000 to 1, and yet the light from the sodium is predominant. This paradoxical situation is partly explained by the low excitation energy required (2.09 e.v.) to produce the *D*-line radiation from the sodium atom. Thus, the fraction of the electrons in the plasma having the 2.09 e.v. required to excite sodium will be greater than the fraction of the electrons possessing sufficient energy to excite neon to the 18-volt levels¹ needed for visible radiation. However, not only must the electrons have the requisite energy, but they must also make collisions with the sodium atoms. If the neon were absent, such encounters would be rare on account of the long mean free path in the low-concentration sodium; but the collisions with the neon atoms actually present cause the electrons to remain in the field longer, and thus the frequency of collisions with sodium atoms becomes appreciable. The neon also provides the main supply of ions for the discharge, although the number of sodium ions increases as the bulb temperature and sodium concentration

¹ Either directly or by cumulative action.

increase. Optimum conditions obtain when the electron energies and sodium concentration are such that copious excitation of sodium atoms results without excessive sodium-ion formation.¹

The high specific luminous efficiency of about 55 lumens/watt, or about three times that of a gas-filled tungsten lamp, has made the use of the sodium lamp attractive in applications where the bright yellow color is not objectionable.

c. High-pressure Hot-cathode Arc Lamps.—The development of mercury-arc lamps² has been in the direction of increased pressures, and has involved the use of an admixed inert gas to facilitate starting and of thermionic cathodes in place of the mercury

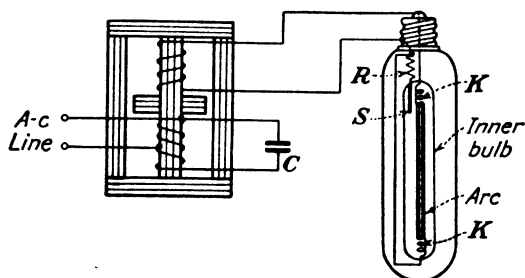


FIG. 11-64.—400-watt high-intensity mercury-arc lamp.

pool of the original Cooper-Hewitt lamps. The resulting improvements have been an increased luminous efficiency, greater brilliancy and therefore a more concentrated light source, and a slight improvement in the color characteristics of the light.

A 400-watt lamp of this type is shown in Fig. 11-64. The inner bulb contains argon gas at a pressure of a few millimeters mercury and a small measured quantity of mercury. When the lamp is switched on, a glow discharge strikes between the starting electrode *S* and the upper oxide-coated cathode *K*, but the discharge quickly changes to an arc between the two cathodes. The cathodes are heated by the discharge itself. At first, the low-pressure argon discharge fills the tube, for the vapor pressure of the mercury is low. As the bulb temperature rises, the mercury vapor pressure increases and the typical blue-green mercury discharge sets in. The amount of mercury is such that

¹ For additional details, see reference 2, cited on p. 450.

² MARDEN, J. W., N. C. BEESE, and G. MEISTER, *Trans. A.I.E.E.*, **55**, 1186, 1936. BUTTOLPH, L. J., *Elec. Eng.*, **55**, 1174, 1936.

it is all vaporized when the pressure reaches approximately 1 atm. At this relatively high pressure the positive column is concentrated down the center of the tube in a narrow pencil. Furthermore, the arc gradient in the positive column rises to the point where the arc voltage is about 150 volts compared with the 20 volts required to maintain the preliminary argon arc. The conditions in the positive column are similar to those discussed in Sec. 11-3.

A high-reactance transformer is used to provide proper starting voltage and to stabilize the arc current. The low power factor of the transformer and lamp (about 0.7) can be increased to 0.94 by the use of a suitable condenser (*C*, Fig. 11-64). Unity power factor cannot be obtained on account of the distortion of the waves. The waves of current and voltage of the arc are similar to those shown in Fig. 11-16.

The specific luminous efficiency of the lamp described above is about 35 lumens per watt. This lamp is subject to a starting time of 5 to 8 min. and to a delay incident to restarting after a power interruption. This delay is necessary to allow the mercury to condense sufficiently so that breakdown can take place.

A group of extra high pressure mercury-arc lamps has been devised in which the arc is confined in a fused quartz tube. In one of the smaller lamps the tube has a bore of 2 mm. and a length of 14 mm. between the oxide-coated¹ electrodes. At the rated power input (85 watts for the small lamp) the mercury pressure builds up to 20 to 30 atm. and the arc concentrates down the center of the tube. Under these conditions the tube wall attains a temperature of 1100°K., and the arc spectrum consists of continuous bands superimposed on the mercury lines. The resulting light has an improved color value. The intrinsic brilliancy is very high, ranging from 18 to 300 cp./sq. mm., and the efficiency ranges from 40 to 50 lumens per watt. In sizes over 250 watts the lamp must be water-cooled.

d. The Fluorescent Lamp.—The fluorescent lamp² is a hot-cathode positive-column type of mercury-arc lamp with the

¹ Recently, improved electrodes comprising metallic thorium supported in a tungsten spiral have been developed. See G. A. Freeman, *Elec. Eng.*, **59**, 445, 1940.

² INMAN, G. E., and R. N. THAYER, *Trans. A.I.E.E.*, **57**, 723, 1938. INMAN, G. E., *Trans. Illum. Eng. Soc.*, **34**, 65, 1939.

inside of the tube coated with a substance that fluoresces under the influence of radiation corresponding to one or more lines of the mercury spectrum. The light of wave length 2,537 Å. in the ultraviolet is especially effective in exciting the fluorescent materials or *phosphors*. In this process a photon of wave length 2,537 Å. is absorbed, and a photon of longer wave length (lesser energy) is emitted.

Phosphors have been found that emit a large share of these secondary photons in the visible range. The emitted light from any single phosphor covers a range of wave lengths with a peak emission that characterizes the color of the light, but by using a mixture of phosphors a color resembling daylight can be obtained. The phosphors consist usually of the metallic silicates, tungstates,

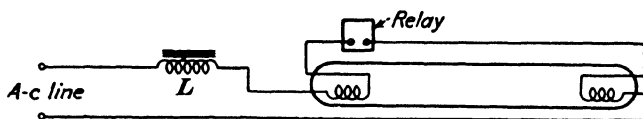


FIG. 11-65.—Fluorescent lamp circuit.

or sulphides, sometimes with a small percentage of added metal "activator." An example is zinc silicate with admixed manganese which fluoresces to give a green color. Some of the phosphors also exhibit phosphorescence; *i.e.*, they emit light for some time after the excitation stops. This property is useful, for it reduces the flicker of the lamps under a.c. operation.

The ultraviolet line 2,537 Å. is produced most abundantly in a low-pressure discharge with relatively low current densities. As a result the tubes are large for a given power consumption; for example, the 15-watt tube is 1 in. in diameter and 18 in. long. One of the circuits used with these lamps is shown in Fig. 11-65. The relay closes the circuit so that the oxide-coated cathodes will be heated, and after a few seconds the relay opens and the discharge starts. Argon gas is utilized in the tube to facilitate starting. The current is stabilized by the choke coil *L* at a value that results in a power factor of about 0.55.

The luminous efficiency of these lamps is about 3 lumens per watt for red, 20 lumens per watt for blue, pink, and gold, 65 lumens per watt for green, and 30 to 35 lumens per watt for "white" and "daylight" lamps. The light output is, however, decreased seriously when the bulb temperature drops much below

30°C. The lamps have a decided advantage over tungsten lamps because of their high efficiency, long life, and lack of glare, but have the drawbacks of high first cost, some flicker and stroboscopic effect, and greater complexity of installation. Undoubtedly, these drawbacks will be minimized and further improvements in efficiency will be realized by continued research and development.

Problems

1. A certain 4-mm. carbon arc has a total voltage drop of 50 volts when the arc current is 16 amp. and a drop of 60 volts when the current is 4 amp. If the arc is operated in a 120-volt d.c. circuit, what load resistance is necessary to give a current of (a) 7 amp. (b) 12 amp.?

2. The arc of Prob. 1 can be described by Eq. (11-4) when the constants have the following values: $a = 38.64$, $b = 2,000$, $c = 10.67$, and $d = 10,670$. If this arc is used, what is the maximum arc length that can be obtained with a circuit voltage of 120 volts direct current and a circuit resistance of 10 ohms? What current would be obtained under this condition?

3. With reference to the arc of Prob. 2, what is the maximum arc length that can be obtained with a circuit voltage of 120 volts and a current of 4 amp.? What value of circuit resistance is necessary for a stable arc under these conditions?

4. What is the maximum resistance that can be used with the arc of Prob. 1, a stable arc still being maintained?

5. When the arc of Prob. 1 is connected in series with a 5-volt-peak 500-cycle alternator, a 120-volt d.c. source of voltage, and a resistance of 16.1 ohms, it behaves as if it consisted of a negative resistance of 4 ohms in series with an inductance of 1.273 millihenrys. Sketch the stable volt-ampere characteristic and the dynamic volt-ampere characteristic.

6. The arc of Prob. 1 is connected in series with a 120-volt d.c. source of voltage and a resistance of 16.1 ohms and is to be used to produce sustained oscillations of 10,000 cycles/sec. Neglecting the inductance of the arc, compute the value of inductance necessary in a parallel circuit when the capacitance used is (a) 0.1 μf , (b) 0.01 μf . What limitation must be placed on the resistance of the parallel branch?

7. A circuit breaker whose contacts separate 1.6 mm. is to be used to interrupt a 500-amp. circuit of nearly pure inductance and having a natural frequency of oscillation of 5,000 cycles/sec. On the assumption that the data of Fig. 11-26 can be applied to the arc, what is the maximum circuit voltage that could be interrupted?

8. Referring to the preceding problem, how many 1.6-mm. arc gaps would be necessary to interrupt the 500 amp. when the circuit voltage is 2,300 volts? Assume that the voltage divides uniformly among the gaps.

9. A certain fiber-tube high-voltage fuse has the reignition characteristic shown in Fig. 11-27. (a) Will this fuse be satisfactory to interrupt a 400-amp. arc in a 2,300-volt circuit when the circuit consists of nearly pure resistance? (b) Will the fuse interrupt a 400-amp. arc in a 2,300-volt circuit when the circuit is highly inductive and has a natural period of 500

cycles/sec.? (c) Will the fuse interrupt a 400-amp. arc in a 1,100-volt circuit when the circuit is highly inductive and has a natural period of 500 cycles/sec.? (d) Explain the results of parts *a*, *b*, and *c*.

10. A two-anode full-wave mercury-arc rectifier is connected by means of an 8-to-1 transformer to a 110-volt source of voltage. If the voltage drop in the rectifier is neglected and the load is assumed to be pure resistance, what is the average d.c. output voltage of the circuit?

11. A certain 15-kw. 300-volt two-anode full-wave mercury-arc rectifier has a constant arc drop of 20 volts when the average load current (resistance load) is varied from 5 to 50 amp. What is the efficiency of the rectifier at (a) full load? (b) Half load? (c) One-quarter load?

12. With reference to the rectifier described in the preceding problem, what would be its efficiency when delivering 50 amp. at (a) 100 volts? (b) 200 volts? (c) 300 volts? (d) 600 volts?

13. A certain mercury-vapor phanotron has rating data as follows:

Shielded filament-type cathode:

Cathode voltage..... 5 volts

Cathode current..... 6.75 amp.

Heating time..... 30 sec.

Maximum peak inverse voltage for condensed mercury temperatures below 60°C. and frequencies less than 150 cycles/sec..... 10,000 volts

Maximum instantaneous anode current at frequencies above 25 cycles/sec..... 5 amp.

Maximum average anode current..... 1.25 amp.

Two of these tubes are used in a single-phase 60-cycle full-wave rectifier circuit supplied by a center-tapped transformer. The rectifier delivers current to a resistance load.

a. Draw the circuit diagram of the rectifier.

b. What is the maximum permissible value of r.m.s. voltage from the center tap to each side of the transformer winding?

c. If the voltage from center tap to each side of the transformer is 2,200 volts r.m.s., what is the smallest permissible value of load resistance?

d. Assume that the rectifier is supplied with 2,200 volts as in *c* and each tube is delivering full rated average current. If the tube drop is constant during the conducting period and has the value 8 volts, what is the efficiency of the rectifier, including the filament loss but transformer losses being neglected?

14. The data for the control characteristic of a gas-filled thyatron are as follows:

V_a , volts	V_g , volts	V_a , volts	V_g , volts
25	0	152	-4
26	-1	225	-4.5
40	-2	337	-5
77	-3	500	-5.5

- a. Plot the control characteristic.
- b. Construct the critical-grid-voltage curve as a function of the time for an anode-circuit alternating voltage of 350 peak volts.
- c. At what instant on the voltage wave will conduction begin if the battery voltage applied in the grid circuit is -4.5 volt?

15. The anode circuit of a thyratron tube contains a resistance load of R ohms and an a.c. source of voltage which may be expressed as $E_m \sin \omega t$. The grid circuit contains a phase-shifting circuit to provide variation of the firing angle ϕ between the zero of the source voltage in the anode circuit and the instant of firing. (a) Show that the expression for the average load current as a function of the firing angle is $(E_m/2\pi R)(1 + \cos \phi)$ when the tube drop is neglected. (b) Discuss the limitations of the formula derived in part a.

16. A mercury-vapor thyratron has rating data as follows:

Indirectly heated cathode	
Heater voltage.....	5 volts
Heater current.....	10 amp.
Heating time.....	300 sec.
Maximum peak inverse anode voltage.....	1,000 volts
Maximum forward anode voltage.....	1,000 volts
Maximum instantaneous control-grid current....	1.0 amp.
Maximum average grid current.....	0.25 amp.
Maximum instantaneous anode current for frequencies over 25 cycles/sec.....	40 amp.
Maximum average anode current.....	6.4 amp.

This tube is to be used in the circuit shown in Fig. 11-52. The transformer furnishes 220 volts r.m.s. at 60 cycles in the anode-to-cathode circuit. Assume that the tube drop may be neglected.

a. Will the tube anode-current ratings be exceeded at any firing angle if the load resistance is 12 ohms?

b. What is the maximum instantaneous anode current if the average anode current is 6.4 amp. and the firing angle takes on the values (1) 30 deg., (2) 60 deg., and (3) 90 deg.?

17. Assume that in Fig. 11-52 the voltage from center tap to each end of the transformer winding is 110 volts and that the frequency is 60 cycles. The condenser C has the value $1.5 \mu f$, and R may be varied from 0 to 5,000 ohms. (a) Over what range of angles can adjustment be made of the phase of the cathode-to-grid voltage relative to that of the anode circuit voltage? Neglect the effect of grid current. (b) Illustrate the voltage relations in this circuit by a vector diagram when R has the value 5,000 ohms.

General References

- 11-1. SLEPIAN, J.: "Conduction of Electricity in Gases," pp. 151-182, Educational Department, Westinghouse Electric & Manufacturing Co., New York, 1933.

- 11-2. THOMSON, J. J., and G. P. THOMSON: "Conduction of Electricity through Gases," Vol. II, Chap. X, Cambridge University Press, London, 1933.
- 11-3. DARROW, K. K.: "Electrical Phenomena in Gases," Chap. XIII, The Williams & Wilkins Company, Baltimore, 1932.
- 11-4. v. ENGEL, A., and M. STEENBECK: "Elektrische Gasentladungen," Vol. II, pp. 119-170, 256-327, Verlag Julius Springer, Berlin, 1934.
- 11-5. DRUYVESTEYN, M. J., and F. M. PENNING: *Rev. Mod. Phys.*, **12**, 140-166, 1940.
- 11-6. LOEB, L. B.: "Fundamental Processes of Electrical Discharge in Gases," pp. 605-639, John Wiley & Sons, Inc., New York, 1939.
- 11-7. WHEATCROFT, E. L. E.: "Gaseous Electrical Conductors," Chaps. VII, X, XI, XII, XIII, Oxford, Clarendon Press, New York, 1938.
- 11-8. DOW, W. G.: "Fundamentals of Engineering Electronics," Chaps. XIX, XXI, XXII, John Wiley & Sons, Inc., New York, 1937.
- 11-9. REICH, H. J.: "Theory and Application of Electron Tubes," Chaps. XI and XII, McGraw-Hill Book Company, Inc., New York, 1939.
- 11-10. MARTI, O. K., and H. WINOGRAD: "Mercury Arc Rectifiers—Theory and Practice," McGraw-Hill Book Company, Inc., New York, 1930.
- 11-11. UYTERHOEVEN, W.: "Elektrische Gasentladungslampen," Verlag Julius Springer, Berlin, 1938.

APPENDIX I

LIST OF SYMBOLS

Symbol	Name of quantity and units	Page where first used
a, a	acceleration (meters per second per second)	46
c	velocity of light (meters per second)	16
c_0	most probable speed (meters per second)	101
$\bar{c}$	arithmetical average speed (meters per second)	107
d_a	dark-space thickness in the abnormal glow	323
d_n	dark-space thickness in the normal glow	322
e	instantaneous electric potential (volts)	174
e	base of Napierian (natural) logarithms	100
e_g	variable component of the instantaneous grid potential (volts)	197
e_p	variable component of the instantaneous plate potential (volts)	197
f, f	force (newtons)	18
f	frequency of voltage source (cycles per second)	309
h	Planck's quantum constant (joule-seconds)	139
i	instantaneous electric current (amperes)	174
i_b	instantaneous plate current (amperes)	195
i_e	electron current (amperes)	341
i_p	positive-ion current (amperes)	341
k	Boltzmann constant (joules per degree)	90
l, s	length (meters)	15
m	mass, mass of a particle (kilograms)	46
m_0	rest mass of electron (kilograms)	63
n	number (of particles, collisions, etc.)	85
p	pressure (newtons per square meter, or millimeters mercury)	82
q, Q	quantity of charge (coulombs)	3
r	radius (meters)	18
s, l	length (meters)	15
t	time (seconds)	46
v	volume (cubic meters)	82
x, y, z	Cartesian coordinates	21
A	area (square meters)	15
A_0	constant in Richardson's equation (amperes per square meter per degree squared)	144
B	magnetic flux density (webers per square meter)	51
C	capacitance (farads)	15

Symbol	Name of quantity and units	Page where first used
C	root-mean-square (r.m.s.) speed (meters per second)	87
D	diffusion coefficient	236
E	static electric potential, potential difference (volts)	6
E_b	actual plate potential (volts)	179
E_B	direct-current (d.c.) circuit voltage, battery voltage	195
E_c	static cathode-to-grid potential (volts)	179
E_c	first critical potential (volts)	260
E_i	ionization potential (volts)	260
E_p	static tube drop (volts)	187
E_s	sparking potential (volts, kilovolts)	288
E_1	difference of potential between the plasma and the probe (volts)	339
F, F	electric intensity (volt per meter)	5
G_p	plate conductance (mhos)	201
G_{pg}	grid-plate transconductance (mhos)	201
H	magnetic intensity (ampere turns per meter)	62
I	electric current (amperes)	6
I_b	static plate current	182
I_c	static grid current	183
J, J	current density (amperes per square meter)	3
J_a	cathode current density in the abnormal glow	323
J_n	cathode current density in the normal glow	323
J_s	temperature-limited thermionic current density	144
K	mobility (meters per second per volt per meter)	225
L	self-inductance (henrys)	204
M_0	molecular weight (kilograms)	83
N	molecular concentration (number per cubic meter)	3
N_0	Avogadro's number (number of molecules per mole)	89
P	power (watts)	198
Q, q	quantity of charge (coulombs)	3
R	resistance (ohms)	6
R	molar gas constant (joules per degree)	83
R_p	plate resistance (ohms)	202
T	absolute temperature (degrees Kelvin)	82
V, V	velocity (meters per second)	46
$\bar{V}$	mean drift velocity (meters per second)	3
V_a	abnormal cathode fall (volts)	323
V_c	cathode fall of potential (volts)	380
V_n	normal cathode fall (volts)	319
V_x	x -component of velocity	50
W	energy, or work (joules)	23
W_a	outer work function	144
W_i	inner work function	139
Z	number of collisions per second	111
α (alpha)	ionizing coefficient for electrons	266
α_r	coefficient of recombination	248

Symbol	Name of quantity and units	Page where first used
β (beta)	ionizing coefficient for positive ions	277
γ (gamma)	coefficient for electron emission due to positive-ion impact on cathode	220
ϵ (epsilon)	permittivity (rationalized m.k.s. system)	15
ζ (zeta)	average number of collisions per meter	230
η (eta)	number of particles (molecules, electrons, etc.) crossing unit area per second	120
θ (theta)	angle (radians)	21
κ (kappa)	dielectric constant	15
λ (lambda)	mean free path (meters)	111
λ	wave length of light	208
λ_e	mean free path of electrons	114
λ_0	long-wave-length limit for photoelectric emission	211
μ (mu)	permeability (rationalized m.k.s. system)	16
μ	amplification factor of grid-controlled tube	179
ν (nu)	frequency of light	209
ν_0	threshold frequency for photoelectric emission	210
ρ (rho)	volume density of charge (coulombs per cubic meter)	33
ρ	resistivity (ohms, meters)	5
σ (sigma)	surface density of charge (coulombs per square meter)	35
σ	molecular diameter	111
τ (tau)	period of rotation	53
ϕ_0 (phi)	net work function (electron volts)	145
ω (omega)	angular velocity (radians per second)	53
Ω (omega)	solid angle (steradians)	30

APPENDIX II

IMPORTANT CONSTANTS

Electronic charge	$q_e = (4.8025 \pm 0.0004) \times 10^{-10}$ e.s.u. $q_e = 1.602 \times 10^{-19}$ coulomb
Electronic rest mass	$m_0 = (9.1070 \pm 0.0014) \times 10^{-31}$ kg.
Velocity of light	$c = (2.99776 \pm 0.00015) \times 10^8$ m./sec.
Planck's constant	$h = (6.6189 \pm 0.0036) \times 10^{-34}$ joule-sec.
Molar gas constant	$R = 8.3136 \pm 0.0010$ joule per degree Kelvin per mole
Avogadro's number	$N_0 = 6.023 \times 10^{23}$ molecules per mole
Boltzmann constant	$k = 1.380 \times 10^{-23}$ joule per degree
Permittivity of free space	$\epsilon = 8.857 \times 10^{-12}$
One standard atmosphere	$= 1.01325 \times 10^5$ newtons/sq. m.
Ice point (absolute scale)	$= 273.18 \pm 0.03^\circ\text{K}.$
Volume of perfect gas at 0°C. and one standard atmosphere	$= (22.4141 \pm 0.0008) \times 10^{-3}$ cu. m. per mole
Napierian base	$e = 2.71828$

NOTE: The values of q_e , m_0 , h , and c are those given by F. G. Dunnington, *Rev. Mod. Phys.*, **11**, 65, 1939. The values of N_0 , k , and ϵ are calculated on the basis of Dunnington's values. The remaining constants are taken from the table of R. T. Birge, *Rev. Mod. Phys.*, **1**, 60, 1929.

APPENDIX III

MOLECULAR WEIGHTS, AND MOLECULAR SPEEDS AND MEAN FREE PATHS AT STANDARD CONDITIONS

(0°C. and one standard atmosphere)

Molecule		Molecular weight M_0 , kg.	R.m.s. speed C , m./sec.	Mean free path λ , m.
H ₂	Hydrogen	2.016×10^{-3}	1838	11.16×10^{-8}
He	Helium	4.002×10^{-3}	1304	17.65×10^{-8}
CH ₄	Methane	16.03×10^{-3}	652	4.89×10^{-8}
NH ₃	Ammonia	17.03×10^{-3}	632	4.28×10^{-8}
H ₂ O	Water	18.02×10^{-3}	615	3.96×10^{-8}
Ne	Neon	20.18×10^{-3}	581	12.53×10^{-8}
N ₂	Nitrogen	28.02×10^{-3}	493	5.96×10^{-8}
C ₂ H ₄	Ethylene	28.03×10^{-3}	493	3.42×10^{-8}
C ₂ H ₆	Ethane	30.05×10^{-3}	476	2.99×10^{-8}
O ₂	Oxygen	32.00×10^{-3}	461	6.44×10^{-8}
HCl	Hydrochloric acid	36.46×10^{-3}	432	4.21×10^{-8}
A	Argon	39.94×10^{-3}	413	6.32×10^{-8}
CO ₂	Carbon dioxide	44.00×10^{-3}	393.6	3.97×10^{-8}
Kr	Krypton	82.9×10^{-3}	286.6	4.85×10^{-8}
Xe	Xenon	130.2×10^{-3}	228.8	3.58×10^{-8}
Hg	Mercury	200.6×10^{-3}	184.2	8.32×10^{-8} *

NOTE: Values in this table are based on the values of E. H. Kennard, "Kinetic Theory of Gases," pp. 24, 148, McGraw-Hill Book Company, Inc., New York, 1938.

* At 219.4°C.

INDEX

Numbers in **bold-face type** indicate pages on which the subject is discussed in some detail.

A

- Abnormal cathode fall, definition, 322
- Abnormal glow discharge, characteristics of, 323
- Accelerating field, effect on work function, 163
- Acceleration of an ion, 46
- Accommodation coefficient, 381*n*
- Air, natural conductivity, 126
- Ambipolar diffusion, 329
- Ambipolar diffusion coefficient, 329
- Amplification factor, of multigrid tubes, 193
 - of a triode, 179
 - effect of structure, 181
 - formulas for, 182
- Amplifier, load voltage wave form of, 197
 - mathematical analysis of, 200
 - operating point of, 196
 - plate current wave form of, 196, 200
 - power consumed, 198
 - power output, 198
 - voltage amplification of, 197, 202
- Amplifier circuit, 194
 - graphical analysis of, 195
 - phototube, 218
- Amplitude control of thyatron rectifier, 436
- Amplitude modulator, 203
- Angular velocity of a charged particle in a magnetic field, 53
- Anode boiling point, effect on arc, 367
- Anode fall in arc, 363
- Anode fall of potential, 344
- Anode in glow, **344**
 - space-charge sheath at, 344
- Anode rays, 71
- Application, of the cathode-ray oscillograph, 65
 - of glow discharge, **348**
 - of grid-controlled tubes, **191**
 - of a high-vacuum diode, 174
 - of phototubes, 213
 - of thyratrons, **436**
- Arc, in air, 7
 - ionizing and deionizing agents in, 133
 - a.c., reignition in, 394, 406
 - anode fall in, 363
 - carbon, 360, 365
 - cathode fall in, 363
 - cold cathode, 383
 - a.c., characteristics, 396
 - reignition in, 395
 - copper, 361, 363
 - effect of boiling point of anode material on, 367
 - energy balance at cathode of, 381
 - gas temperature in, 365
 - gold, 361
 - hissing, 367
 - hot cathode, 383
 - ignition of, 362
 - ignition of, 358
 - ion concentration in, 371
 - ion density in, 364
 - mercury, 361
 - positive column of, convection loss from, 372
 - diffusion in, 372

- Arc**, positive column of, energy losses from, 372
 recombination in, 372
 radiation from, 372
 short, reignition voltage for, 406
 temperature of gas in, 365, 370
 thermal conduction from, 372
 thermionic emission in, 380
 transition from glow, 358, 380
 discontinuous, 360
 for low-melting point electrodes, 361
 tungsten, 360
 cathode in, 359
 positive-ion current at cathode of, 382
 turbulent, 402
 vacuum, 364
 volt-ampere characteristics, 366
- Arc cathode**, 379
 effect of impurities on, 385
 high-field emission theory of, 384
- Arc characteristics, a.c.**, 393
 carbon, 365
 dynamic, 366
 dynamic a.c., 393
 total, 363
- Arc cross section**, 374
- Arc discharge**, definition of, 357
 formation of, 287
- Arc drop in mercury phanotron**, 429
- Arc duration**, in circuit interruption, 404
- Arc extinction**, 397
 theories of, in oil circuit breakers, 413
 turbulent, theory of, 401
- Arc in fiber tube**, 401
- Arc hysteresis**, 387
- Arc ignition in a thyatron**, 362
- Arc interruption, a.c.**, 397, 399
 d.c., 398
- Arc lamps**, hot cathode, high pressure, 454
 low pressure, 452
- Arc oscillator**, 388
 Poulsen, 392
- Arc positive column**, 364, 368
- Arc radius**, 374
- Arc in small tube**, 378
 deionization in, 401
 extinction of, 400
- Arc stability**, 368, 389
- Arc-back**, in mercury-arc rectifiers, 415, 420
 effect of pressure on, 424
 probability of occurrence, 423
 theory of, 422
 in phanotron, 429
- Arithmetical average molecular speed**, 107
- Arrester**, lightning (*see* Lightning arrester)
- Atoms**, gas, ionization by collision of electrons with, 265
 ionization by collision of positive ions with, 268
 metastable (*see* Metastable atoms)
- Attachment constant**, 247
- Attachment of electrons to molecules**, 247
- Avogadro number**, 89
- Avogadro's hypothesis**, 88
- Ayrton's formula**, 366
- B**
- Backfire in mercury-arc rectifier**, 421
- Beam tube**, 193
- Boiling point of anode material**, effect on arc, 367
- Boltzmann equation**, 172
- Boltzmann's gas constant**, 90
- Breakdown**, 287
- Breakdown, time lag of**, 296
- Breakdown characteristic of thyatron**, 431
- Breakdown in a thyatron**, 432
- Breakdown voltage**, 290
- Brush discharge**, 290
- C**
- Carbon arc**, 360, 365
- Cathode**, indirectly heated, 158

- Cathode, oxide coated, 155
 activation of, 156
 emission constants of, 157
 thermionic, cathode sputtering of, 428
 heat shielded, 158
 positive-ion bombardment of, 160
 thoriated tungsten, activation of, of, 151
 production of, 150
 unipotential, 158
- Cathode of arc, **379**
 effect of impurities on, 385
 energy balance at, 381
 high-field emission theory of, 384
- Cathode dark space, 312, **316**
 electric intensity in, 317
 measurement of thickness of, 316
 theory of, 318
- Cathode efficiency, 159
- Cathode fall, dependence on work function in glow discharge, 319
 normal and abnormal, definitions of, 322
 normal, values of, 321
- Cathode fall in arc, 364
- Cathode fall in glow, effect of gas mixtures, 320
- Cathode glow, 313
- Cathode of phanotron, ratio of electron current to positive-ion current, 428
 space-charge sheath at, 427
- Cathode-ray oscillograph, **54**
 applications of, 65
 for frequency determination, 66
 electrostatic deflection type, **54**, 57, 59
 high voltage, 56
 magnetic deflection type, **60**
 sweep-voltage wave for, 67
 types of, 55
- Cathode spot, 379, 383
- Cathode sputtering of thermionic cathodes, 428
- Cathode in tungsten arc, 359
 positive-ion current at, 382
- Cesiated cathodes, thermionic emission from, 154
- Characteristic curves for a triode, 183
- Charge carriers, 125
- Charge of an ion, 3
- Child's law, 170
- Circuit breaker, Deion, **409**
 Deion air, 406, 410
 oil blast, 408
 operating time, 403
 reignition voltage in, 406
- Circuit breaker arc extinction, diffusion theory of, 413
 displacement theory of, 413
- Circuit breaker classification, **407**
- Circuit breakers, **403**
 high voltage, 408
 low voltage, 412
 oil, 409
 theories of arc extinction in, 413
- Circuit interrupters, **403**
- Circuit interruption, arc duration in, 404
 effect of circuit on, 404
 problem of, 403
- Coefficient of diffusion, ambipolar, 329
- Cold-cathode a.c. arc, characteristics, 396
 reignition in, 395
- Cold-cathode arc, 383
- Cold-cathode electron emission, 165
- Collision, fraction of energy of an ion lost in, 229
 theory of ionization by, 271
- Collision frequency, 111
- Collisions, ionizing, number of, per meter, 266
 between molecules, 93, 109
 made by an ion while advancing unit distance, 230
 of positive ions with gas atoms, ionization by, **268**
- Color temperature, 214

470 THEORY OF GASEOUS CONDUCTION AND ELECTRONICS

- Column, positive (*see* Positive column)
 - Composite surfaces, electron emission from, **149**
 - Concentration, electron, in a metal, 138
 - ionic, 3
 - of ions in arc, 371
 - molecular, 89
 - Concentric cylinders, field between, 34
 - Condensed mercury temperature, 429
 - Conductance, plate, 201
 - of high-vacuum amplifier tube, 201
 - Conduction, electronic, definition of, 2
 - gaseous, definition of, 2
 - mechanism of, 2
 - metallic and gaseous, 5
 - Conductivity, natural, of air, 126
 - of a gas, **125**
 - Constants, physical, 464
 - Contact difference of potential, **160**
 - relation to work function, 161
 - Continuous-arc oscillator, 390
 - Control characteristic of thyatron, 431, 433
 - Convection loss from arc positive column, 372
 - Copper arc, 361, 363
 - Corona, between concentric cylinders, **305**
 - loss on transmission lines, **308**
 - in nonuniform fields, **301**
 - on parallel wires, **305**
 - Corona discharge, 288, 301, 305
 - theory of, 307
 - Corona voltage, 290
 - for parallel wires, **306**
 - Cosmic rays, as an ionizing agent, 126
 - Coulomb, 18
 - Coulomb's law, 17
 - Critical grid voltage for thyatron rectifier, 437
 - Critical *pd* product, 294
 - Critical potential, definition of, 262
 - values of, 260
 - Critical voltage, disruptive, 309
 - Crookes dark space, 312, **316**
 - Cross section of arc, 374
 - Current, effect on the positive-column gradient, 331
 - space-charge-limited, 138
 - Current density, of electrons crossing a potential barrier, 340
 - in a gas, definition of, 2
 - normal, in glow, 322
 - in Townsend discharge, 273, 277
 - Current in a phototube, dependence on pressure, 283
 - Current-voltage characteristics, effect on stability, 345
 - Cyclotron, 73
 - Cylinders, concentric, corona between, **305**
 - sparkling between, **305**
- D
- Dalton's law of partial pressures, 91
 - Dark discharge, 287
 - Dark space, cathode (*see* Cathode dark space)
 - Dark space, Crookes, 312, **316**
 - Dark space, Faraday, 313, **325**
 - Dark space thickness, normal, 322
 - Degenerate gas, 139
 - Deion air circuit breaker, 406, 410
 - Deion circuit breaker, 409
 - Deionization in arc in small tube, 401
 - Deionization time in thyatron, 435
 - Deionizing agents, definition of, 126
 - diffusion, 130, **235**
 - electric field, 130, **225**
 - recombination in the gas, 130, **244**
 - recombination at surfaces, 131, **254**
 - Diameter, molecular, 111
 - Dielectric constant, 15
 - relative, 18

- Diffusion, 130, **235**
 - ambipolar, 329
 - in arc positive column, 372
 - deionizing agent, 130
 - importance in positive column of glow, 328
 - ionizing agent, 128
 - number of molecules diffusing across unit surface area, 236
 - between parallel electrodes with uniform ionization, 241
 - Diffusion coefficient, 236
 - ambipolar, 329
 - dependence on pressure and temperature, 241
 - derivation of, 238
 - experimental values of, 240
 - Diffusion theory of circuit breaker arc extinction, 413
 - Diode, cylindrical, space-charge-limited electron current in, 173
 - gas filled (*see* Phanotron)
 - high vacuum, applications of, 174
 - peak voltmeter, 176
 - potential minimum in, 137, 171
 - potentials and intensities in, 136
 - rectifier, 174
 - saturation in, 136
 - space charge in, 136
 - space-charge-limited electron current in a parallel plane, 167
 - tube, 135
 - voltmeter, 176
 - Direct current power transmission, 446
 - Discharge, arc (*see* Arc discharge)
 - brush, 290
 - corona, 288, 301, 305
 - theory of, 307
 - dark, 287
 - glow (*see* Glow discharge)
 - initiation of, in phanotron, 426
 - non-self-maintaining, 270, 278
 - self-maintaining, 278
 - conditions for, 279
 - stability of, **345**
 - Discharge, Townsend, **270**
 - current density in, 273, 277
 - secondary emission in, 275
 - tube, neon, 451
 - types of, **287**
 - Disintegration voltage, 428
 - Displacement theory of circuit breaker arc extinction, 413
 - Disruptive critical voltage, 309
 - Distribution of free paths, **116**
 - Distribution functions, general forms of, **96**
 - Distribution law, Fermi-Dirac, for components of velocities, 141
 - for total velocities, 139
 - Maxwell-Boltzmann, for molecular speeds, 103
 - for components of velocities, 104
 - Maxwellian, compared with Fermi-Dirac, 141
 - Distribution laws, velocity, Maxwell's deduction of, **97**
 - Maxwell-Boltzmann, representation of, **102**
 - Double-grid thyatron, 435
 - Double sheath, 427
 - Drift, velocity for an ion, 225
 - Dynamic a.c. arc characteristics, 393
 - Dynamic arc characteristics, **386**
 - Dynamic resistance, 8, 331, 391
 - Dynatron region of tetrode characteristic, 192
- E**
- Electric field, as a deionizing agent, 130, **225**
 - energy gained by an ion in, 47, 49
 - forces in, 45
 - fringing of, 57
 - as an ionizing agent, 127
 - mapping of, **37**
 - by the electrolytic method, **39**
 - motion of ions in, 46, 50
 - nonuniform, sparking and corona in, **301**

- Electric field, strong, electron emission in, 162, **165**
 ion mobility in, 229, 232
 uniform, sparking intensity in, **296**
 sparking potential in, **290**
 velocity of an ion in, 50
 weak, ion mobility in, 226, 232
- Electric intensity, **19**
 in cathode dark space, 317
 between concentric cylinders, 34
 outside a conducting surface, 35
 in a glow discharge, 313
 line of, 19, 37
 between parallel plates, 36
 and potential, relation between, **27**
 surface integral of, 29
 in a uniform space-charge region, **33**
- Electric potential, **23**
- Electrodes and walls, recombination on, **254**
- Electron, **1**
 average life of, 247
 energy near a metal, 146
 longitudinal mass of, 65
 mean free path of, 114
 rest mass of, 63
 transverse mass of, 65
 ultimate energy of, 280
- Electron attachment to form negative ions, 246
- Electron avalanche, 307
- Electron beams, focusing of (*see* Focusing of electron beams)
- Electron concentration, determination of, in a plasma, 342
 in a metal, 138
- Electron current, ratio to positive-ion current at cathode in phanotron, 428
 space-charge-limited, 138, **167**
- Electron emission, calculation by Fermi theory, 144
 cold cathode, 129, 162, 165
 from composite surfaces, **149**
 effect of temperature on, 137
 photoelectric, 129, **206**
- Electron emission, in presence of accelerating field, 164
 saturation current density, 144
 secondary, caused by electron impact, **186**
 in a strong electric field, 129, 162, **165**
 temperature limited, 138, **147**
 thermionic, **135**
 from cesiated cathodes, 154
 Electron emitter, oxide coated, **155**
 thoriated tungsten, 149
- Electron impacts, number per second against gas molecules, 247
- Electron lens, 70
- Electron multiplier, 190
- Electron space current, grid control of, **177**
- Electron temperature, 340
- Electron volt, 49
- Electron yields, secondary, for electron impact, 188
- Electronic conduction, definition of, **2**
- Electronics, scope of, **1**
- Electrons, crossing a potential barrier, 340
 Fermi-Dirac distribution law for components of velocities of, 141
 free, 5
 ionization by collision of, with gas atoms, **265**
 in metals, theory of, **138**
 and positive ions, recombination of, in a gas, 246
- Electrostatic deflection type of cathode-ray oscillograph, **54, 57, 59**
- Electrostatic field, definition of, 17
- Electrostatic focusing of electron beams, 70
- Emission, high field, 129, 162, **165**
 theory of arc cathode, 384
 photoelectric, 129, **206**
 spectral distribution curve of, **208, 214**

- Emission, secondary, 129
 thermionic, 129
- Emission constants, of oxide-coated cathodes, 157
- Emitters, practical electron, 149
 thoriated tungsten, 149
 thermionic constants for atomic film, 150
- Energy, of electron near a metal, 146
 gained by an ion in a field, 47, 49
 of ion, fraction lost in a collision, 229
 of ion, terminal, 230
 for ionization, 259
 liberated by an ion coming up to a wall, 255
 losses from arc positive column, 372
 molecular, and temperature, 90
 of secondary electrons, 188
 of a single molecule, 90
 ultimate, of an electron, 280
- Energy balance at cathode of arc, 381
- Energy level diagram, mercury, 262
- Energy levels, 261
- Equipotential surface, 37
- Equipotential surfaces in positive column of glow, 329
- Excited states, 261
- Extinction, of arcs, 397
 arc, in oil circuit breakers, theories of, 413
 of arc in a small tube, 400
 theory of, for a turbulent arc, 401
- F**
- Faraday dark space, 313, 325
- Fermi theory, calculation of electron emission by, 144
- Fermi-Dirac distribution law for components of velocities, 141
- Fermi-Dirac velocity distribution law, 139
- Fiber tube, arc in, 401
- Field emission of electrons, 129, 162, 165
- Floating potential, 339
- Fluorescent lamp, 455
- Focusing of electron beams, electrostatic, 70
 gas, 69
 magnetic, 69
- Force, in an electric field, 45
 on an ion, 17, 46
 in a magnetic field, 51
- Forward voltage of thyatron, 436
- Fraction of energy of ion lost in a collision, 229
- Free path, mean (*see* Mean free path)
- Free paths, 109
 distribution of, 116
- Frequency, of arc oscillator, 391
 collision, 111
 determination by the cathode-ray oscillograph, 66
 threshold, 259
- Fringing of the electrostatic field, 57
- Fuse, high-voltage, 407
- G**
- Gas, conductivity of, 125
 degenerate, 139
 kinetic picture of, 84
 temperature in arc, 365, 370
- Gas constant, Boltzmann's, 90
- molar, 83
- Gas focusing of electron beams, 69
- Gas mixtures, effect on cathode fall of glow, 320
- Gaseous conduction, definition of, 2
 scope of, 1
- Gaseous discharge lamps, 450
- Gaseous and metallic conduction, 5
- Gases, general equation for, 82
 mixed, 91
 pressure in, 91
- Gauss's theorem, 29, 32
- Glow, cathode, 313
 negative (*see* Negative glow)
- Glow discharge, abnormal, characteristics of, 323
 anode, 344

Glow discharge, appearance, 311,

316

applications, **348**

definition of, 311

dependence of cathode fall on
work function in, 319effect of gas mixtures on cathode
fall in, 320electric intensity distribution in,
313equipotential surfaces in the posi-
tive column of, 329

first cathode layer in, 316 •

formation of, 288, 315

importance of diffusion in positive
column of, 328

mechanism of maintenance, 318

normal current density in, 322

normal, use in lightning arresters,
352

potential distribution in, 313

space-charge sheath at anode in,
344

volt-ampere characteristics of, 314

Glow-to-arc transition, 358, 380

discontinuous, 360

for low-melting-point electrodes,
361

Glow-tube lamp, 451

Glow-tube oscillator, **348**

continuous, 350

intermittent, 349

Gold arc, 361

Gradient, positive column (*see* Posi-
tive column, gradient)Gravitational analogy, **48**Grid, control of electron space
current, **177**

in high-power rectifier, 424

screen, 191

suppressor, 193 •

of thyatron, function of, 431

Grid-controlled tubes, applications
of, **191**Grid-current characteristic of a
triode, 184

of thyatron, 434

Grid-plate transconductance, 201

H

High-field emission, 129

theory of arc cathode, 384

High-voltage generator, cyclotron,
73

Van der Graaf type, 71

Hissing arc, 367

Hot-cathode arc lamps, high pres-
sure, 454

low pressure, 452

Hysteresis in the arc, 387

I

Iconoscope, 76

Igniter, 444

Ignition, of arc, 358

in a thyatron, 362

in hot-cathode arc tube, 362

Ignitron, **443**

rectifier, 445

welding control, 445

Image dissector, 74

Image force and work function,
146Image method, Kelvin's, **40**Impurities, effect on arc cathode,
385in gas, effect on cathode fall
of glow, 320

Intermittent arc oscillator, 389

Interphase transformer, 417

Interrupters, circuit, **403**

Interruption, a.c. arc, 397, 399

circuit, arc duration in, 404

effect of circuit on, 404

problem of, 403

d.c. arc, 398

Intensities and potentials in a diode,
136Intensity, sparking, in a uniform
field, **296**Inverter, **446**

thyatron, 447

Ion, acceleration of, 46

angular velocity in a magnetic
field, 53

- Ion, charge of, 3**
 collisions made while advancing unit distance, 230
 coming up to wall, energy liberated by, 255
 concentration in arc, 371
 definition of, 1
 density, in arc, 364
 in negative glow, 326
 in positive column, 328
 energy of, 47, 49
 force on, 17, 46
 in a magnetic field, 51
 fraction of energy lost in a collision, 229
 mass of, 1
 mobility in a strong field, 229, 232
 in a weak field, 226, 232
 molecular, 1
 motion in electric field, 46, 50
 in magnetic field, 51
 negative, electron attachment to form, 246
 path in a magnetic field, 52
 terminal energy, 230
 velocity of drift of, 225
 velocity in an electrostatic field, 50
Ion paths, radius of curvature in a magnetic field, 52
Ionic concentration, 3
Ions, positive, ionization by collisions of, with gas atoms, 268
 radial drift of, 372
Ionization, by collision, 127
 of electrons with gas atoms, 265
 of positive ions with gas atoms, 268
 theory of, 271
 with a volume source of ionization, 274
 by flames, chemical effects, etc., 128
 in gas, photoelectric, 259
 multiple, 268
 by successive impacts, 266, 364
 thermal, 128, 263
 Ionization energies, 259
 Ionization potential, definition of, 259
 second, definition of, 261
 values of, 260
 values of, 260
 Ionization time, in thyratron, 435
 Ionizing agents, cosmic rays, 126
 definition of, 126
 diffusion, 128
 electric field, 127
 flames, chemical effects, etc., 128
 high-field emission, 129
 ionization by collision, 127
 photoelectric emission, 129
 radioactivity, 127
 secondary emission, 129
 thermal ionization, 128
 thermionic emission, 129
 X rays, 127
 Ionizing coefficient, α , 266, 279, 280, 282, 283
 β , 277
 γ , 276, 284
 Ionizing collisions per meter, number of, 266
 Ionizing and deionizing agents, in an arc in air, 133
 examples of action of, 131
 in a gas-filled phototube, 131
 in a gas-filled rectifier tube, 131
 in a high-vacuum phototube, 132
 in a neon glow discharge, 133
 Ionizing efficiency, 267
 Ionizing process, 1
 J
 Joule, the, 14
 K
 Kelvin's method of images, 40
 Kinetic picture of a gas, 84
 Konel alloy, 156

L

- Lamps, gaseous-discharge, **450**
- Life, average, of electron in a gas, **247**
- Light, measurement and units of, **216*n***
- Light intensity, effect on photo-electric emission, **212**
- Lightning arrester, **351**
 - porous block, **378**
 - use of normal-glow discharge in, **352**
- Line of electric intensity, **19, 37**
- Lissajous figures, **66**
- Load line, **196**
- Load voltage wave form of amplifier, **197**
- Longitudinal mass of an electron, **65**

M

- Magnetic-deflection type cathode-ray oscillograph, **60**
- Magnetic field, angular velocity of a charged particle in, **53**
 - force on an ion in, **51**
 - motions of ions in, **51**
 - path of an ion in, **52**
 - radius of curvature of an ion path in, **52**
- Magnetic focusing of electron beams, **69**
- Mapping of electric field by the electrolytic method, **39**
- Mapping of electric fields, **37**
- Mass, longitudinal, of an electron, **65**
 - of oxygen molecule, **93**
 - rest, of an electron, **63**
 - transverse, of an electron, **65**
- Mass correction, relativity, **62**
- Maxwell-Boltzmann velocity distribution laws, representation of, **102**
- Maxwellian distribution law compared with Fermi-Dirac law, **141**
- Maxwell's velocity distribution laws, deduction of, **97**
- Mean free path, **110**
 - comparison with experiment, **117**
 - effect of temperature and pressure on, **112**
 - electronic, **114**
 - of various gases, **465**
- Mercury, energy level diagram of, **262**
- Mercury arc, **361**
- Mercury-arc lamp, **369, 454**
- Mercury-arc rectifier, **414**
 - arc-backs in, **415, 420**
 - probability of occurrence, **423**
 - backfire in, **421**
 - efficiency, **418**
 - high power, grids in, **424**
 - shields in, **424**
 - historical development, **10**
 - operation of, **416**
 - polyphase, **417**
 - pressure in, **415**
 - theory of arc-back in, **422**
 - two anode, **414**
 - wave forms in, **416**
- Mercury temperature, condensed, **429**
- Mercury-vapor pressure data, **415**
- Mercury-vapor rectifier, hot cathode, **7**
- Metallic and gaseous conduction, **5**
- Metastable atoms, **262**
 - effect on sparking potential, **299**
 - effect on striations, **336**
- M.k.s. system of units, **14**
- Mobility, **225**
 - calculation of, in a strong field, **229**
 - in a weak field, **226**
 - comparison with experiment, **233**
 - dependence on pressure, **233**
 - general expression for, **231**
 - ionic, in a strong field, **229, 232**
 - in a weak field, **226, 232**
 - methods for determining, **234**
 - Thomson formula, **229**
- Modulator, amplitude, **203**
- Molar gas constant, **83**

Molecular collisions, 93, 109
 Molecular concentration, 89
 Molecular diameter, 111
 Molecular energy and temperature, 90
 Molecular speed, 88
 and pressure, 85
 Molecular speeds, arithmetical average, 107
 list of, 465
 Maxwell-Boltzmann, distribution law for, 103
 r.m.s. average, 106
 Molecular theory of gases, 81
 Molecular velocities, distribution of, 93
 Molecular weight, 83
 list of, 465
 Molecule, energy of, 90
 oxygen, mass of, 93
 Molecules, crossing a surface, number of, 118
 gas, number of electron impacts per second against, 247
 number of, diffusing across unit surface area, 236
 Motion of ions, in an electrostatic field, 46, 50
 in a magnetic field, 51
 Multigrid thermionic tubes, 191
 Multigrid tubes, amplification factor of, 193
 Multiple ionization, 268

N

Negative glow, 313, 325
 ion density in, 326
 Neon discharge tube, 451
 ionizing and deionizing agents in, 133
 Newton, the, 14
 Non-self-maintaining discharge, 270, 278
 Nonuniform fields, sparking and corona in, 301
 Normal cathode fall, definition of, 322

Normal cathode fall, values of, 321
 Normal current density in the glow, 322
 Normal dark space thickness, 322
 Normal glow discharge, use in lighting arresters, 352

O

Ohm's law, 6
 Oil blast circuit breaker, 408
 Oil circuit breakers, 409
 Operating point, stable, 347
 of an amplifier, 196
 Operating time of circuit breaker, 403
 Oscillations, plasma, 336
 Oscillator, arc, 388
 continuous, 390
 frequency of, 391
 intermittent, 389
 Poulsen, 392
 glow tube, 348
 continuous, 350
 intermittent, 349
 vacuum tube, 203
 Oscillator circuit, wave forms in, 205
 Oscillograph, cathode ray (*see* Cathode-ray oscillograph)
 Oscilloscope, 56
 Oxide-coated cathodes, preparation of, 155
 Oxygen molecule, mass of, 93

P

Parallel plates, field between, 36
 Partial pressures, Dalton's law of, 91
 Paschen's law, 290, 293
 Path of an ion in a magnetic field, 52
 Pentode tube, 192
 Perfect gas, 82
 Permeability, 14
 Permittivity, 15, 18
 relative, 18
 Phanotron, 424
 arc ignition in, 362

- Phanotron, arc-back in, 429
 cathode of, ratio of electron current to positive-ion current, 428
 initiation of discharge in, 426
 mercury, arc drop in, 429
 potential distribution in, 426
 space-charge sheath at cathode of, 427
 tube drop in, 425, 428
- Phase-controlled thyatron rectifier, 439
- Phosphors, 456
- Photoelectric electron emission, 129, 206
 effect of light intensity on, 212
 net work function, 210
 quantum yield in, 212
 spectral distribution curve of, 208, 214
 temperature effect on, 211
 theory of, 209
 threshold frequency for, 210
- Photoelectric ionization in the gas, 259
- Photoelectric tubes (*see* Phototube)
- Photon, 210
- Phototube, 213
 amplifier circuit for, 218
 application of, 213
 dependence of current on pressure, 283
 development of, 9
 gas filled, 7, 270
 high vacuum, 6
 ionizing and deionizing agents in, 132
 preparation of, 213
 sensitivity rating of, 216
 time lag in, 219
 vacuum and gas filled, 215
- Physical constants, 464
- Plasma, 336, 338
 determination of electron concentration in, 342
 oscillations, 336
 probe measurements in, 337
 velocity distributions in, 342
- Plate characteristic, 183
- Plate conductance, 201
- Plate current of a triode, 182
- Plate current wave form of an amplifier, 196, 200
- Plate resistance, 203
- Poisson's equation, one-dimensional, 168
- Polyphase mercury-arc rectifier, 417
- Porous-block lightning arrester, 378
- Positive column, 313, 327
 of an arc, 364, 368
 convection loss from, 372
 diffusion in, 372
 energy losses from, 372
 recombination in, 372
 characteristics, determination of, 337
 of a glow, equipotential surfaces in, 329
 importance of diffusion in, 328
 recombination in, 314
 gradient, effect of current on, 331
 of pressure on, 332
 of size of discharge tube on, 331
 ion density in, 328
 striated, 334
 temperature of, 333
 uniform, 327
- Positive ions, and electrons, recombination of, in a gas, 246
 ionization by collision of, with gas atoms, 268
- Positive-grid thyatron, 434
- Positive-ion bombardment, secondary electron yields for, 220
 of a thermionic cathode, 160
- Positive-ion current, at cathode of tungsten arc, 382
 ratio of, to electron current at cathode of phanotron, 428
- Positive-ion impact, secondary electron emission due to, 220
- Positron, 1
- Potential, anode fall of, 344
 contact, 160
 critical, definition of, 262
 values of, 260

- Potential, floating, 339
 ionization, definition of, 259
 values of, 260
 minimum, in a diode, 137, 171
 second ionization, values of, 260
 space, 341
 sparking (*see* Sparking potential)
 unit of, 24
- Potential barrier, current density of
 electrons crossing, 340
- Potential difference, 26
- Potential distribution, in a glow
 discharge, 313
 in a phanotron, 426
- Potential and electric intensity,
 relation between, 27
- Potential increase, 26
- Potential at a point, 23
- Potentials and intensities in a
 diode, 136
- Poulsen arc oscillator, 392
- Power transmission by d.c., 446
- Pressure, effect on arc-backs, 424
 effect on the current in a photo-
 tube, 283
 on diffusion coefficient, 241
 on the ionizing coefficient α ,
 280, 283
 on mobility, 233
 on the positive column gradient
 332
 on sparking potential, 293
 in a mercury-arc rectifier, 415
 in mixed gases, 91
 and molecular speed, 85
 and temperature, effect on the
 mean free path, 112
 on recombination coefficient,
 254
- Pressure data for mercury vapor, 415
- Principle of superposition, 20, 25
- Probe measurements in a plasma,
 337
- Probes, sheaths on, 338

Q

- Quantum yield in photoelectric
 effect, 212

R

- Radial drift of ions, 372
- Radiation from arc, 372
- Radioactivity, as an ionizing agent,
 127
- Radius, of arc, 374
 of curvature of ion paths in a
 magnetic field, 52
- Recombination, in arc positive col-
 umn, 372
 coefficient, 248
 calculation of, 250
 effect of temperature and pres-
 sure on, 254
 method for measuring, 249
 values of, 250
 of electrons and positive ions in a
 gas, 246
 in the gas, 130, 244
 in the positive column of the
 glow, 314
- Langevin theory, 253
 at surfaces, 131, 254
- Thomson theory, 251
 on walls and electrodes, 254
- Rectifier, diode, high vacuum, 174
 half wave, wave form of current
 in, 175
 hot cathode, mercury vapor, 7
 ignitron, 445
 mercury arc (*see* Mercury-arc rec-
 tifier)
 thyatron (*see* Thyatron rectifier)
- Rectifier tube, gas filled, ionizing
 and deionizing agents in, 131
- Reignition, in an a.c. arc, 394, 406
 in cold-cathode a.c. arcs, 395
 in turbulent arcs, 406
- Reignition voltage, in circuit break-
 ers, 405
 definition of, 394
 for a short arc, 406
- Relativity mass correction, 62
- Resistance, 7
 dynamic, 8, 331, 391
 line, 347
 plate, 203

Resistivity, 5
 Rest mass of an electron, 63
 Richardson line, 147
 Richardson's classical theory for thermionic emission, 120
 R.m.s. average molecular speed, 107

S

Saha equation, 264
 Saturable reactor, 442
 Saturation current density for electron emission, 144
 Saturation in a diode, 136
 Schottky effect, 162
 Screen grid, 191
 Secondary electron emission, 129
 caused by electron impact, 186
 coefficient, γ , 276, 284
 due to positive-ion impact, 220
 in a Townsend discharge, 275
 Secondary electron yields, for electron impact, 188.
 for positive-ion bombardment, 220
 Secondary electrons, energy of, 188
 from insulating surfaces, 191
 Secondary emission (*see* Secondary electron emission)
 Self-maintaining discharge, 278
 conditions for, 279
 Sheath, space charge, at phanotron cathode, 427
 Sheaths on probes, 338
 Shield-grid thyatron, 435
 Shields in high-power rectifier, 424
 Similitude, principle of, 304
 Sodium vapor lamp, 453
 Space charge, in a diode, 136
 electric intensity in a spherical region of, 33
 in a triode, 178
 Space-charge sheath, at anode in glow, 344
 at phanotron cathode, 427
 Space-charge-limited electron current, 138, 167
 in a cylindrical diode, 173
 Space-charge-limited electron current, effect of initial velocities of electrons on, 170
 in a parallel-plane diode, 167
 Space potential, 341
 Sparking, between concentric cylinders, 305
 criterion for, 292
 in nonuniform fields, 301
 theory of, 291
 Sparking intensity, in a uniform field, 296
 Sparking potential, 288
 dependence on temperature and pressure, 293
 effect of metastable atoms on, 299
 in high-pressure gases, 298
 minimum, 294
 in mixed gases, 298
 in a uniform field, 290
 in vacuum, 298
 values of, 294
 Spark-over, 287
 Spectral distribution curve, of photoelectric emission, 208, 214
 for tungsten lamp, 214
 Speed, most probable, 102
 Sputtering at a thermionic cathode, 428
 Stable operating point, 347
 Stability, arc, 368, 389
 of discharges, 345
 effect of voltage-current characteristics on, 345
 Striated positive column, 334
 Striations, effect of metastable atoms on, 336
 Successive impacts, ionization by, 266, 364
 Sun lamp, type S-1, 360
 Supercontrol tube, 194
 Superposition, principle of, 20, 25
 Suppressor grid, 193
 Surface integral of electric intensity, 29
 Sweep voltage for the cathode-ray oscillograph, 67

Switch, vacuum, **353**

Symbols, list of, 461

T

Television, **74**

receiving tubes, **78**

Temperature, color, 214

condensed mercury, 429

effect on diffusion coefficient, 241

on electron emission, 137

on photoelectric emission, 211

on sparking potential, 293

electron, 340

of gas in an arc, 365, 370

and molecular energy, **90**

of positive column, 333

and pressure, effect on mean free path, 112

effect on recombination coefficient, 254

Temperature-limited electron emission, 138, **147**

Terminal energy of an ion, 230

Tetrode, 191

characteristic, dynatron region of, 192

Theories of arc extinction in oil circuit breakers, 413

Theory, of arc-back in mercury-arc rectifier, 422

of cathode dark space, 318

of corona discharge, 307

high-field emission, of arc cathode, 384

of ionization by collision, 271

molecular, of gases, **81**

for recombination, Langevin's, 253

Thomson's, 251

Richardson's classical, for thermionic emission, **120**

of sparking, 291

of turbulent arc extinction, 401

Thermal conduction from arc, 372

Thermal ionization, 128, **263**

Thermionic arc, 383

Thermionic cathode, heat shielded, 158

positive-ion bombardment of, 160

sputtering of, 428

Thermionic constants, for atomic-film emitters, 150

for clean metals, 148

Thermionic emission, 129

in arc, 380

from cesiated cathodes, 154

introduction, **135**

Richardson's classical theory of, **120**

theory of, **138**

Thermionic tubes, multigrid, **191**

Thermionic vacuum tube, development of, **9**

Thoriated tungsten cathode, activation of, 151

production of, 150

Thoriated tungsten electron emitter, 149

Threshold frequency, 259

for photoelectric emission, 210

Thyratron, **431**

applications, **436**

arc ignition in, 362

breakdown in, 432

breakdown characteristic of, 431

control characteristic of, 431, 433

deionization time in, 435

double-grid type, 435

forward voltage rating of, 436

function of grid of, 431

grid-current characteristics of, 434

ignition time in, 435

positive-grid type, 434

shield-grid type, 435

Thyratron inverter, 447

Thyratron rectifier, 436

amplitude control of, 436

critical grid voltage for, 437

phase-controlled, 439

wave forms in, 438

Thyratron welding control, 439

Time lag of breakdown, 296

- Townsend discharge, **270**
current density in, **273, 277**
secondary emission in, **275**
- Transconductance, grid plate, **201**
- Transfer characteristic, **184**
- Transformer, interphase, **417**
- Transition from glow to arc, **358, 380**
discontinuous, **360**
for low-melting-point electrodes, **361**
- Transmission lines, corona loss on, **308**
- Transmission of power by d.c., **446**
- Transverse mass of an electron, **65**
- Triode, amplification factor of, **179**
effect of structure on, **181**
formulas for, **182**
characteristic curves for, **183**
gas filled (*see* Thyatron)
grid currents in, **184**
high vacuum, **177**
electron paths in, **181**
plate current of, **182**
space charge in, **178**
- Triode voltmeter, **206**
- Tube, amplifier, power consumed in, **198**
beam, **193**
oscillator, **203**
pentode, **192**
photoelectric, **213**
development of, **9**
supercontrol, **194**
television receiving, **78**
tetrode, **191**
thermionic, development of, **9**
multigrid, **191**
type S-1 sun lamp, **360**
variable μ , **194**
- Tube conductance, **201**
- Tube drop in phanotron, **425, 428**
- Tungsten arc, **360**
cathode in, **359**
positive-ion current at cathode of, **382**
- Tungsten lamp, spectral distribution curve for, **214**
- Turbulence theory of arc extinction, **401**
- Turbulent arc, **402**
reignition in, **406**
- U**
- Uniform positive column, **327**
- Units, fundamental, **12**
fundamental derived, **13**
m.k.s. system of, **14**
rationalization of, **15**
unified systems of, **13**
- V**
- Vacuum arc, **364**
- Vacuum switch, **353**
- Vacuum tube, thermionic, development of, **9**
- Van der Graaf high-voltage generator, **71**
- Variable- μ tube, **194**
- Velocities, molecular, distribution of, introduction, **93**
- Velocity components, Maxwell-Boltzmann law for distribution of, **104**
- Velocity distribution law, Fermi-Dirac, **139**
Maxwell's deduction of, **97**
Maxwell-Boltzmann, for components of velocities, **104**
for molecular speeds, **103**
representation of, **102**
- Velocity distribution in a plasma, **342**
- Velocity of drift of an ion, **225**
- Velocity of an ion in an electric field, **50**
- Visual sensitivity curve, **214**
- Volt, definition of, **24**
- Volt-ampere characteristic, arc, **366**
of a glow discharge, **314**
- Voltage, breakdown, **290**
corona, **290**
for parallel wires, **306**
disruptive critical, **309**

Voltage, peak inverse, 174
Voltage amplification of an amplifier, 197, 202
Voltage-current characteristics, effect on stability, 345
Voltmeter, diode, 176
slide-back, 177
triode, 206

W

Wall, energy liberated by an ion coming up to, 255
Walls and electrodes, recombination on, 254
Wave form, of current in half-wave rectifier, 175
load voltage, of an amplifier, 197
in a mercury-arc rectifier, 416
in an oscillator circuit, 205
plate current, of an amplifier, 196, 200
of sweep voltage for the cathode-ray oscillograph, 67
Wave form, in a thyatron rectifier, 438
Welding control, 439, 445
ignitron, 445
Wires, parallel, corona on, 305
parallel, corona voltage for, 306
Work function, classical, 121
effect of accelerating field on, 163
inner, 145
net, 145
for photoelectric emission, 210
outer, 145
reduction by an ion layer, 153
relation to cathode fall in a glow discharge, 319
relation to contact difference of potential, 161
values of, 260
by photoelectric tests, 211
values for clean metals, 148
Work function and image force, 146

X

X rays, as an ionizing agent, 127

